No evidence for globally coherent warm and cold periods over the pre-industrial Common Era

Raphael Neukom[1*], Nathan Steiger[2], Juan José Gómez-Navarro[3], Jianghao Wang[4] & Johannes P. Werner[5]

May 13, 2019

1. Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, Switzerland;
2. Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA;
3. Department of Physics, University of Murcia, Spain;
4. The MathWorks, Inc., Natick, USA;
5. Bjerknes Center for Climate Research, Bergen, Norway
*corresponding author: Raphael Neukom (neukom@giub.unibe.ch)

Abstract

Earth's climate history is often understood by breaking it into constituent climatic epochs[1]. Over the Common Era (the past 2,000 years) these epochs, such as the "Little Ice Age"[2–4],have been characterized as temporally coherent across extensive spatial scales[5]. While the rapid global warming seen in observations over the past 150 years shows nearly global coherence[6], the spatio-temporal coherence of climate epochs over the Common Era has yet to be robustly tested. Here we find, based on global paleoclimate reconstructions over the past 2,000 years, no evidence for pre-industrial globally-coherent cold and warm epochs. In particular, the coldest multi-decadal to centennial epoch of the last millennium, the putative Little Ice Age, had the highest probability of occurring during the 15^{th} century over most of the central and eastern Pacific, during the 17^{th} century in northwestern Europe and southeastern North America, and during the mid-19^{th} century over most remaining regions. Furthermore, the spatial coherence that does exist over the pre-industrial Common Era is consistent with the spatial coherence of stochastic climatic variability. This lack of spatio-temporal coherence indicates that pre-industrial forcing was not sufficient to produce globally synchronous extreme temperatures at multi-decadal and centennial timescales. In contrast, we find that the warmest period of the last two millennia occurred simultaneously in the 20^{th} century for over 98% of the globe. This provides strong evidence that anthropogenic global warming is not only unparalleled in terms of absolute temperatures[5], but also unprecedented in spatial consistency within the context of the last 2,000 years.

The study of past climate provides a critical baseline from which to understand and contextualize changes in the contemporary climate. Since the formative period of modern Earth sciences in the 1800s, the complex history of Earth's climate has been conceptualized by the construction of distinct climatic periods or epochs[1,7]. Within the era of the past 2,000 years, several terms for climatic epochs have come into wide use. Most prominent among these is the "Little Ice Age" (LIA), a term which was originally created to broadly describe late Holocene glacier growth in the Sierra Nevada mountains[2]; later the LIA was used to describe inferred late Holocene glacial

advances in many locations, particularly the European Alps[3,4]. Over the past few decades, this term has been very widely used in paleoclimatology and historical climatology to indicate a nearly global, centuries long cold climate state that occurred roughly between 1300 and 1850 CE (Common Era)[5,8]. This period is often contrasted with the Medieval Warm Period, also known as the Medieval Climate Anomaly (MCA)[8–10] and is commonly associated with warm temperatures between the years 800-1200 CE. The first millennium of the Common Era has also been subdivided into the "Dark Ages Cold Period" (DACP)[11,12], or "Late Antique Little Ice Age" (LALIA)[13], situated within about 400-800 CE and lastly the "Roman Warm Period" (RWP)[12,14], which covers the first few centuries of the Common Era. We note that for all of these epochs, no consensus exists about their precise temporal extent.

Each of these climatic epochs has its origin in pieces of paleoclimatic evidence from the extratropical Northern Hemisphere, particularly Europe and North America[4,9–12]. Climate epoch narratives were constructed to explain the early paleoclimatic evidence and subsequently developed time series from across the globe were situated within these narrative frameworks. This process likely created the expectation that climate epochs were global-scale phenomena. Loosely defined epochs based on a few dozen specific proxies were hard to falsify given the inherent noise of natural proxies, with, for example, nearly all annually resolved proxies covering the Common Era having a signal-to-noise ratio of less than 1, and usually below 0.5 (Ref. 15). Yet the association of a relatively small number of paleoclimate proxy records with global-scale phenomena did not come without controversy and the discovery of proxy time series that did not match the standard climatic epoch narratives[4,10,16]. Studies that have attempted to assess the spatial coherence of Common Era climate epochs have used relatively few proxy records (e.g., 14 proxy time series[17]), have used only continentally averaged temperature reconstructions[18], or have used only one or two reconstruction methods[8,12], a choice which has been shown to limit the reliability of the assessment of temperature patterns[19].

Here we test the hypothesis that there were globally coherent climate epochs over the Common Era using a collection of probabilistic, global temperature reconstructions over 1-2000 CE derived from a set of six different ensemble field reconstruction methodologies (Methods; note that we use "coherence" here in its general, non-signal-processing sense). The reconstructions are based on techniques that vary widely in their assumptions and approaches to the reconstruction problem. They span a broad range of complexity using

basic proxy composites at the one end and advanced statistical techniques at the other that incorporate physical constraints and forcing information via climate model simulations. All methods use a common input dataset, the annual records from the recent PAGES 2k global temperature-sensitive proxy collection[20] (Fig. **1**; Methods). This multi-method, probabilistic framework allows us to robustly assess the spatiotemporal homogeneity of climatic variability over the Common Era.

At the native annual resolution, the reconstruction ensemble mean shows no clear indication of a long period of years with globally consistent below-average temperatures relative to the mean of 1-2000 CE (Fig. **2**a); the area fraction of warmth and cold shows high inter-annual variability. 97% of the years prior to 1850 had at least 10% of the globe experiencing above average temperatures and 10% of the globe experiencing below average temperatures. It is only if the reconstructed time series are smoothed over multi-decadal time scales (Methods) and if global area is shown in aggregate that the classical picture of a loosely defined LIA and MCA appears (Fig. **2**b and Extended Data Fig. **1**). Yet, the analysis in Fig. **2** does not include information from individual ensemble members (Extended data Fig. **1**), nor does it indicate the spatial patterns of coherence, or provide a precise evaluation of the climate epochs hypothesis.

To quantify the spatial coherence of cold and warm epochs, we consider the time of occurrence of a climate anomaly as the variable to be characterized within a probabilistic framework. We calculate the most probable period of peak warming or cooling during each of the five climatic epochs discussed previously (Methods). At each grid point location, we identify the warmest 51-year average within the epochs commonly referred to as warm: RWP, MCA and current warm period (CWP). Analogously, we identify the coldest 51-year average for the DACP, and LIA cold epochs. Given the lack of objective definitions for these epochs, we keep a wide window for the search of peak warming or cooling for each period (Methods and Fig. **3**). Note that to assess the CWP, we search for the warmest peak within the entire Common Era. The century within which we find the highest ensemble probability for maximum warming or cooling at each location is shown in Fig. **3**.

There is considerable spatial heterogeneity in the timing of temperature maxima and minima (Fig. **3**). No pre-industrial epoch shows global coherence in the timing of the coldest or warmest periods. There is, however, regional coherence. For example, there are almost continental-scale patterns during many of the periods and there is a coherent pattern in the tropical Pacific

in the RWP, DACP and LIA periods, reminiscent of the El Niño–Southern Oscillation, the most dominant mode of inter-annual variability in the climate system[21].

In contrast to the spatial heterogeneity of the pre-industrial era, the highest probability for peak warming over the entire CE (Fig. **3**c) is found in the late 20$^{\text{th}}$ century almost everywhere (98% of global surface area) except for Antarctica, where contemporary warming has not yet been observed over the entire continent[22]. Thus even though the recent warming rates are not entirely homogeneous over the globe with isolated areas showing little warming or even cooling[22,23], the climate system is now in a state of global temperature coherence unprecedented over the Common Era.

Through a boostrapping uncertainty analysis (Methods, Extended Data Fig. **2**), we find that the particular spatial patterns shown in Fig. **3** are robust. Furthermore, the heterogeneity in the timing of maxima and minima is an inherent property of the input proxy data, which show a similar lack of global coherence in the timing of each putative climate epoch (Extended Data Fig. **3**).

Is the amount of spatial coherence in the pre-industrial period consistent with stochastic climate variability? We find that it is (Fig. **4**) and that the spatial agreement across all reconstructions is typically low: In 84% of reconstruction ensemble members, less than 50% of the global area fraction agrees on the timing of the warmest or coldest 51-year peaks across all pre-industrial epochs (Fig. **4**). This supports the results shown in Fig. **3**, providing evidence that peak pre-industrial warm and cool periods occurred at different times at different locations. In contrast, the CWP shows distinct temporal and spatial agreement with the warmest multi-decadal peak of the Common Era occurring in the late 20$^{\text{th}}$ century. The area fraction agreeing on the timing of the CWP is significantly larger than that expected from stochastic climate variability (Mann-Whitney U test, $p < 0.01$, $n_1 = n_2 = 600$).

In addition to using an unprecedented collection of reconstruction methods to test the climate epochs hypothesis, we conducted a range of sensitivity experiments, including noise proxy reconstructions (Methods). These confirm the robustness of our results to the specific proxy network, the reconstruction parameter choices, potential biases arising from the selection and calibration of proxies over the observational period, and the specific statistical tests of spatio-temporal coherence (see Methods, and Extended Data Figs **4**-**7**). In addition, we confirmed the lack of pre-industrial spatial coherence in last millennium climate model simulations (Extended Data Fig. **8**).

And as in the reconstructions, the spatial consistency seen in model simulations over the 20$^{\text{th}}$ century suggests that anthropogenic global warming is the cause of increased spatial temperature coherence relative to prior eras.

An important caveat to our results is that the spatio-temporal distribution of high resolution proxy data is inherently unequal and often sparse. Future improvements in this regard may lead to better resolved spatial patterns, especially in the Southern Hemisphere and during the first millennium, where uncertainties in our reconstructions are highest (Fig. **1** and Extended Data Figs **9-10**). Though such improvements are unlikely to lead to greater global coherence when the extant proxy data do not show indications of such (Extended Data Fig. **3**).

The results shown here can explain at least two curious facts about climate epochs of the Common Era: the lack of consensus about the timing of climate epochs and the discovery of records that do not fit the standard narratives. Peak warming and cooling events appear to be regionally constrained. Anomalous globally averaged temperatures during certain periods do not imply the existence of epochs of globally coherent and synchronous climate. This global asynchronicity suggests that multi-decadal regional extremes are driven by regionally specific mechanisms, being either unforced internal climate variability[24,25] or regionally varying responses to external forcings[26–28].

Given these results, we advocate for a regional framing for understanding climate variability of the pre-industrial Common Era. Likewise, the interpretation of individual paleoclimate time series should not be force-fit into global narratives or epochs. Rather, the *a priori* belief about a given paleoclimate time series should be that it represents local information, with the extent of its correlation length scale to be justified and not assumed. In this framing, specific records can provide regional tests of the mechanisms of climate variability[29,30] while collections of many records can address larger scales. Against this regional framing, perhaps the most striking result shown here is the exceptional spatio-temporal coherence during the warming of the 20$^{\text{th}}$ century. This result provides further evidence of the unprecedented nature of anthropogenic global warming in the context of the last 2,000 years.

References

1. Köppen, W. & Wegener, A. *Die Klimate der geologischen Vorzeit* (Gebrüder Borntraeger, Berlin, 1924).
2. Matthes, F. E. Report of Committee on Glaciers, April 1939. *Eos, Transactions American Geophysical Union* **20,** 518–523 (1939).
3. Grove, J.M. *The Little Ice Age* (Methuen, London, 1988).
4. Matthews, J. A. & Briffa, K. R. The 'little ice age': re-evaluation of an evolving concept. *Geografiska Annaler: Series A, Physical Geography* **87,** 17–36 (2005).
5. Masson-Delmotte, V. *et al.* in *Climate Change 2013 - The Physical Science Basis* (eds Intergovernmental Panel on Climate Change & Intergovernmental Panel on Climate Change) 383–464 (Cambridge University Press, Cambridge, 2013). ISBN: 978-1-107-66182-0.
6. IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* 1535 pp. ISBN: ISBN 978-1-107-66182-0 (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013).
7. Brückner, E. *Klimaschwankungen seit 1700 nebst Bemerkungen über die Klimaschwankungen der Diluvialzeit* (E. D. Hölzel, Wien and Olmütz, 1890).
8. Mann, M. E. *et al.* Global Signatures and Dynamical Origins of the Little Ice Age and Medieval Climate Anomaly. *Science* **326,** 1256–1260 (2009).
9. Lamb, H. H. The early medieval warm epoch and its sequel. *Palaeogeography, Palaeoclimatology, Palaeoecology* **1,** 13–37 (1965).
10. Bradley, R. S., Hughes, M. K. & Diaz, H. F. Climate in Medieval time. *Science* **302,** 404–405 (2003).
11. Helama, S., Jones, P. D. & Briffa, K. R. Dark Ages Cold Period: A literature review and directions for future research. *The Holocene* **27** (2017).

12. Ljungqvist, F. C. A New Reconstruction of Temperature Variability in the Extra-Tropical Northern Hemisphere During the Last Two Millennia. *Geografiska Annaler: Series A, Physical Geography* **92,** 339–351 (2010).

13. Büntgen, U. *et al.* Cooling and societal change during the Late Antique Little Ice Age from 536 to around 660 AD. *Nature Geoscience* **9,** 231–236 (2016).

14. Röthlisberger, F. *10,000 Jahre Gletschergeschichte der Erde* (Sauerländer, Aarau, 1986).

15. Wang, J., Emile-Geay, J., Guillot, D., Smerdon, J. E. & Rajaratnam, B. Evaluating climate field reconstruction techniques using improved emulations of real-world conditions. *Climate of the Past* **10,** 1–19 (2014).

16. Bradley, R. 1000 Years of Climate Change. *Science* **288,** 1353–1355 (2000).

17. Osborn, T. J. The Spatial Extent of 20th-Century Warmth in the Context of the Past 1200 Years. *Science* **311,** 841–844 (2006).

18. PAGES2k Consortium. Continental-scale temperature variability during the past two millennia. *Nature Geoscience* **6,** 339–346 (2013).

19. Wang, J., Emile-Geay, J., Guillot, D., McKay, N. P. & Rajaratnam, B. Fragility of reconstructed temperature patterns over the Common Era: Implications for model evaluation. *Geophysical Research Letters* **42,** 7162–7170 (2015).

20. PAGES2k Consortium. A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data* **4,** sdata201788 (2017).

21. McPhaden, M. J., Zebiak, S. E. & Glantz, M. H. ENSO as an Integrating Concept in Earth Science. *Science* **314,** 1740–1745 (2006).

22. Stenni, B. *et al.* Antarctic climate variability on regional and continental scales over the last 2000 years. *Clim. Past* **13,** 1609–1634 (2017).

23. Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G. & Saba, V. Observed fingerprint of a weakening Atlantic Ocean overturning circulation. *Nature* **556,** 191–196 (2018).

24. Wang, J. *et al.* Internal and external forcing of multidecadal Atlantic climate variability over the past 1,200 years. *Nature Geoscience* **10,** 512–517 (2017).

25. Delworth, T. L. *et al.* The North Atlantic Oscillation as a driver of rapid climate change in the Northern Hemisphere. *Nature Geoscience* **9,** 509–512 (2016).

26. Hegerl, G. C., Brönnimann Stefan, Schurer Andrew & Cowan Tim. The early 20th century warming: Anomalies, causes, and consequences. *Wiley Interdisciplinary Reviews: Climate Change* **9,** e522 (2018).

27. Abram, N. J. *et al.* Early onset of industrial-era warming across the oceans and continents. *Nature* **536,** 411–418 (2016).

28. Bindoff, N. L. *et al.* in *Climate Change 2013 - The Physical Science Basis* (eds Intergovernmental Panel on Climate Change & Intergovernmental Panel on Climate Change) 867–952 (Cambridge University Press, Cambridge, 2013). ISBN: 978-1-107-41532-4.

29. Collins, M. *et al.* Challenges and opportunities for improved understanding of regional climate dynamics. *Nature Climate Change* **8,** 101 (2018).

30. Xie, S.-P. *et al.* Towards predictive understanding of regional climate change. *Nature Climate Change* **5,** 921–930 (2015).

Acknowledgements

This is a contribution to the PAGES 2k initiative. PAGES 2k consortium members are acknowledged for providing input proxy data. Julien Emile-Geay provided the graphEM-infilled temperature target grid. Some calculations were run on the Ubelix cluster of the University of Bern. R.N. is supported by the Swiss NSF grant PZ00P2_154802. N.S. was supported by the NOAA Climate and Global Change Postdoctoral Fellowship Program administered by UCAR's Visiting Scientist Programs and by the U.S. National Science Foundation OISE-1743738 and AGS-1805490. LDEO contribution number XXXX. J.J.G.N. acknowledges the "Juan de la Cierva-Incorporación" program (grant IJCI-2015-26914), as well as the CARM for the funding provided through the Seneca Foundation (projects 20022/SF/16 and 20640/JLI/18).

Author contributions

The study was conceived by R.N., N.S. and J.J.G.N. with inputs from all authors. Reconstructions were performed by J.J.G.N.(AM reconstruction method), R.N. (CPS, PCR, CCA), N.S. (DA), J.W. (GraphEM and earlier version of CCA). R.N. ran the analysis of reconstruction results and made the figures. J.P.W ran validation analyses. N.S. and R.N. wrote the manuscript with discussion and contribution from all authors.

Author information

Competing Interests

The authors declare that they have no competing interests.

Correspondence

Correspondence and requests for materials should be addressed to neukom@giub.unibe.ch

Figure legends

Figure 1 | Spatiotemporal proxy data availability. **a** Map of the proxy records[20] used for the reconstructions by archive type. Shading indicates the distance of a given $5° \times 5°$ grid cell to the closest proxy record(s). **b** Temporal availability of proxy data for each archive type, coded by color as in **a**. The red line (right y-axis) shows the width of the 90% confidence interval of the unfiltered reconstructions, latitude weighted, and averaged over all methods. Values are relative to the instrumental temperature standard deviation over 1911-1995 CE.

Figure 2 | Distribution of warm and cold temperatures over the Common Era. Global area fractions with warm (red shading) and cold (blue shading) temperature anomalies with respect to a 1-2000 CE reference period (see Methods). Shading intensity indicates the magnitude of warmth and cold. **a** Annual unfiltered data. **b** 51-year butterworth filtered data.

Figure 3 | Timing of peak warm and cold periods. Century with the highest ensemble probability of containing the warmest (**a,b,c**) and coldest (**d,e**) 51-year period within each putative climatic epoch (see Methods). The full range over which the search was performed for each epoch is indicated in brackets in the title of each map.

Figure 4 | Spatial consistency of warm and cold periods. The proportion of Earth's surface (vertical-axis) that simultaneously experienced the warmest (top) or coldest (bottom) multi-decadal period (51 years) during each of five different epochs (see Methods). Each solid circle is an ensemble member plotted according to the year (horizontal-axis) in which the largest area experienced peak warm/cold conditions. Gray shading represents the distribution from the same analysis based on multivariate AR1 noise fields, with darker colors indicating higher probability. Boxplots on the right show area fractions integrated over time (center line, median; ends of boxes, interquartile range; whiskers, 90% range). Bold text indicates epochs with reconstructed area fractions that are significantly higher than those of the noise fields (Mann-Whitney U test, α=0.05); only for the CWP do the area fractions from the reconstructions exceed those of the noise fields. This figure shows that the peak warming and cooling varies temporally across ensemble members (circles distributed over a range of dates on the horizontal-axis) and only affects a limited fraction of global surface area (low absolute values on the vertical-axis) prior to the industrial era.

Methods

Instrumental target

We use the HadCRUT4 global temperature grid[31] on $5^\circ \times 5^\circ$ spatial resolution infilled using GraphEM to have a complete global coverage over the calibration period[20]. We use annual values aggregated over the April-March seasonal window, which reflects the 'tropical year' and in contrast to a calendar year window, does not interrupt the austral growing season.

Proxy data

The data from the PAGES 2k temperature database v2.0.0 (Ref. 20) are used as predictors. For the results shown in the main text, we use a screened network based on regional false-discovery-rate screening (R-FDR)[20], which reduces the number of proxies from originally 688 to 257. The spatial coverages of the full and screened networks are displayed in Extended Data Fig. **3** and Fig. **1**, respectively. The screened network yields improved reconstruction skill for most methods over much of the globe. However, our conclusions are robust to the choice of either the full or screened network (Extended Data Fig. **4**). The current implementation of some of the reconstruction methods

used herein does not allow us to incorporate proxy data with gaps, or missing values over the calibration period. Therefore we only use records with annual or higher resolution (210 records; Fig. **1**), thereby using mainly records with very small to negligible age uncertainties (85% tree or coral archives). Records of subannual resolution are averaged to annual resolution over the April to March window. Missing values in the proxy matrix over the calibration/validation window (2.2%) were infilled using DINEOF[32].

Reconstruction parameters

The calibration period used for all methods is 1911-1995 CE. This reflects a compromise between using as many years as possible to cover a large temperature range for calibration and using a period with sufficient spatial coverage of instrumental data (at the beginning of the calibration period) or proxy records (at the end). 100-member ensembles are used for all methods to generate the analyses and plots presented in the manuscript, yielding a total ensemble size of 600.

The period 1881-1910 CE is used for validation to compare the relative performance of the reconstruction methods (Extended Data Figs **9** and **10**). We use the following metrics to assess the performance of our reconstructions: The continuous ranked probability score[33,34] (CRPS), which has been conceptually adapted to mimic the Reduction of Error (RE) and Coefficient of Efficiency (CE) scores[35]. The CRPS of the reconstructions is subtracted from the CRPS generated from surrogates based on the instrumental target data following Refs 36 and 37. CRPS_RE (CRPS_CE) compares the mean potential CRPS of the reconstruction with the instrumental surrogates over the 1911-1995 CE calibration (1881-1910 CE validation) period. In contrast to the traditional RE and CE measures, these metrics are strictly proper scoring rules[33]. For a detailed description of the metrics see Ref. 37. The root mean squared error (RMSE) of the ensemble median over the validation period is divided by the instrumental standard deviation at each location to allow a relative comparison at different locations. Finally we calculate the Pearson correlation coefficient of the reconstruction ensemble median with the target over the validation period.

While the spatial distribution of the proxy network is global, the Northern Hemisphere contains more proxies than the Southern Hemisphere and the number of proxies decreases back in time (Fig. **1**). Consequently, the reconstructions generally have highest confidence and inter-method agree-

ment closer to the present, (Fig. **1**b and Extended Data Fig. **10**), and nearest to the proxy locations (Extended Data Fig. **9**). We note that no particular method stands out as being particularly more or less skillful than any other (Extended Data Fig. **9**).

Reconstruction methods

Composite Plus Scale (CPS)

CPS is a widely used index reconstruction method that has been used to reconstruct local to global mean climate[38,39]. The input proxy data are averaged into a composite time series, which is then scaled to the mean and standard deviation of the reconstruction target over the calibration period. Here we use a point-by-point[40] implementation of CPS, probably the most simple approach to reconstruct a climate field. This method does not make use of the spatial covariance structure of the target temperature field, which may lead to unrealistic spatial consistency in the reconstructed fields. On the other hand, the point-by-point approach is, in contrast to most other CFR techniques, not bound to the problematic assumption that the spatial temperature patterns in the calibration period are stable over time.

Here, we use the CPS implementation of Ref. 41, which weights the proxy record by their correlation with the (grid cell) target. We do not limit the proxies used at each location by using a maximum search radius, as the spatial decorrelation distance is not uniform over the globe and to allow information from teleconnected areas to be included in the reconstructions. We use an ensemble approach similar to Ref. 41 combining uncertainties arising from parameter decisions and calibration errors. The following reconstruction parameters are resampled for each reconstruction member: Proxy selection (removing 10% of records), calibration period (a block of 10 years within the 1911-1995 CE calibration period is removed), proxy weight (the correlation-based weight is multiplied by a factor within 1/1.5 and 1.5.). Ensemble perturbation based on the calibration error is implemented as in Ref. 15: Multivariate noise time series with the same standard deviation as the residuals between the target and reconstructed field over the calibration period are added to each ensemble member. We use first order autoregressive (AR(1)) noise with the same AR(1) coefficients as the residuals. We use a nested approach, which means that the reconstruction process is repeated for each time period over the CE with unique proxy availability. The results

of each nest are spliced together to obtain a reconstruction covering the full CE.

Principal Component Regression (PCR)

PCR has been widely used in climate field reconstructions[42–46]. This method reduces the dimensions of both the target field and the proxy matrix using principal component (PC) analysis. The instrumental PCs are predicted back in time based on regression with the proxy PCs and then back-transformed to the spatial dimensions of the target grid using the loadings from the PCA. As such, this approach assumes that the covariance structure of the temperature grid remains the same over the reconstruction period as in the time window used for calibration. Here, we use the PCR approach introduced by Ref. 42 and an ensemble integration similar to Refs. 41 and 44. The nested and ensemble approach is identical to CPS described above with the following exceptions: The random weight factor is multiplied with the weight of each proxy derived from the PCA. We also resample the PC truncation parameters for the proxy (instrumental) matrices in a way that the retained PCs explain between 40% and 90% (60% and 99%) of the total variance. We resample this parameter for two reasons: because there are multiple existing approaches to truncating PCs without an objectively discernible best method; and because the truncations are often sensitive to the period over which the PCA is performed.

In contrast to earlier studies using PCR we do not re-adjust the variance of the reconstructed field to the target variance over the calibration period. This re-adjustment is often done to avoid strong variance changes between the reconstructions of the different proxy nests. We do not apply this correction here because in our case, the differences in variance between the adjusted and non-adjusted reconstructions are relatively small. The largest effect of the re-adjustment is that it leads to a large ratio of low- vs. high frequency variance in the reconstructions, particularly over the high northern latitudes, leading to reconstructed temperatures that are very low compared to the other methods over the entire reconstruction period.

Canonical Correlation Analyis (CCA)

CCA employs singular value decomposition (SVD) to perform dimensional reductions separately on instrumental temperature matrix, proxy matrix, and

the regression coefficient matrix which describes their relationships[47,48]. The basic assumption, as in most paleoclimate applications, is that the first few leading modes of EOF-PC pairs contain most of the variance in the target climate field and the multiproxy network. The algorithm seeks an optimal set of truncation parameters that yields good approximations of the above-mentioned matrices. These truncation parameters are chosen by minimizing the area-weighted root mean square error (RMSE) of the reconstruction relative to the target field using a leave-half-out cross-validation procedure. Specifically, a separate set of parameters was obtained for each proxy nest, so that the algorithm accounts for the heterogeneity in data availability and thus can adaptively regularize the regression matrix. This adaptive procedure was developed in a recent study[15,19] as an improvement to the original algorithm[48]. Ensemble perturbations were done as described for PCR above.

GraphEM

GraphEM[49] is based on the theory of Gaussian graphical models (GGMs or Markov random fields). A GGM makes use of the conditional independence structure of the climate field, in order to reduce the dimensionality and obtain a parsimonious estimate of the inverse of the covariance matrix $\hat{\Sigma}$. The conditional independence relations are estimated by solving an l_1-penalized maximum likelihood problem[50]. Σ is then estimated in accordance with these conditional independence relations. The resulting $\hat{\Sigma}$ is sparse and better conditioned, and therefore is applicable within the Ordinary Least Squares framework. This procedure is implemented within the standard expectation maximization (EM) algorithm without further need for regularization.

Specifically, the covariance model is chosen via the graphical lasso algorithm[50]. Three sparsity parameters need to be specified to determine the graphical structure, separately for the temperature-temperature (TT), proxy-proxy (PP), and the temperature-proxy (TP) parts of the covariance matrix. The values need to be large enough that the true graph is contained in the estimated one, but small enough for the covariance matrix to be well-conditioned. For this study, the parameters have been set to (TT, TP) = (2%, 2%), with a diagonal matrix for PP, reflecting the conditional independence of proxies given the temperature field.

Data assimilation (DA)

We use an off-line DA technique that optimally combines proxy data with climate model states[51,52]. The climate model provides an initial, or prior, state estimate that is updated based on the proxy observations and an estimate of the errors in both the observations and the prior. For the annual reconstructions in this work, the reconstruction is made by iteratively computing the state update equations of DA for each year of the existing proxy data without the need to run an ensemble of online simulations (see Ref. 53 for precise mathematical details and links to DA paleoclimate reconstruction code). As in previous work[53], we construct the prior using simulation number 10 from the Community Earth System Model Last Millennium Ensemble (CESM LME)[54].

For the prior we specifically used the middle 998 years of the CESM LME simulation excluding the two simulation endpoints to create a static 998 member prior ensemble that was used to estimate the climate state in each year of the reconstruction. As in Ref. 53, we performed sensitivity tests with different members from the CESM LME and found no discernible differences in the results.

As in Ref. 52, the proxies in the DA framework are modeled as linear univariate responders to temperature; the errors in the model's estimate of the proxies are taken as the residuals of a local linear univariate fit between the proxies and HadCRUT4 over the calibration period. This DA method naturally provides uncertainty estimates for the target field in the form of an ensemble of equally likely state estimates for each year. For the analyses performed here, we used a random sampling of 100 of these state estimates from the available 998 members.

Analog Method (AM)

The AM method has been successfully used as a CFR technique to reconstruct global temperature fields[55]. The method requires the existence of a pool of plausible climate fields, which can be obtained from computer simulations, observations or combinations of both. As in Ref. 55, we use the ensemble of available simulations within the PMIP3 project, which consists of 18,327 annual temperature fields. For direct comparison with the climate fields, each proxy time series is converted into a "local temperature reconstruction" through a local linear univariate fit with HadCRUT4. We then

calculate the distance between each field and the local, proxy-based reconstructions in a given year; the distance is based on minimizing the spatial RMSE between the local reconstructions and the pool sampled at the closest grid point to each proxy location. The full CFR is then the average of the 5 fields that most closely resemble the local reconstructions at a given time step[55]. Thus the spatial structure of temperature is provided by the different climate models, while the temporal evolution is obtained from the information contained in the network of proxies.

To produce an ensemble that accounts for the probabilistic nature of this reconstruction and their uncertainties, we apply a bootstrap-based approach. We sub-sample both the pool of analogs and the local reconstructions so that in each instance only half of the information is used. This degradation of the available information naturally leads to spread within the ensemble that is larger around locations where the pool is loosely constrained by the proxies, and therefore allows us to quantify the uncertainties implicit in the reconstruction.

Spatial anomalies (Fig. 2)

To generate this figure, the ensemble median reconstructions of each method and at each grid cell are first centered to the reference period 1-2000 CE. The (area weighted) fraction of grid cells exceeding the temperature thresholds displayed in the legend margin of Fig. **2** is then calculated. The values from the six methods are then averaged as displayed in Fig. **2**. In panel b, the reconstruction time series are smoothed with a 51-year butterworth lowpass filter prior to the analysis.

Cold and warm period analysis

Figure 3

To calculate the timing of peak warming/cooling over the climatic epochs, we first calculate 51-year running temperature averages for each location and ensemble member. For each warm (cold) epoch and grid-cell, the year with the maximum (minimum) value is then identified, using the center-year of the warmest (coldest) 51-year period. In the following we use the term "peak-year" for this maximum (minimum). We then identify the century, within which the largest number of members (over the combined 600-member

reconstruction ensemble) have the peak-year. This century is then indicated in Fig. **3** for each grid point. The maps thus show the century, within which the multi-decadal peak warming (or largest cooling) is most probable for each epoch. The DACP and LIA minima are searched within the first and second Millennium CE, respectively. RWP maxima are allowed to occur within 1-750 CE, MCA maxima between 751-1350 CE, and the CWP is assessed in the context of the full 2,000 years of the Common Era.

Uncertainty analysis for Figure 3

Uncertainties in the analysis are quantified by bootstrapping. We recalculated peak warm/cold analysis described above 1000 times using 600 bootstrap samples drawn from the reconstruction ensemble members (Extended Data Fig. **2**). We find that the particular spatial patterns shown in Fig. **3** are robust, with at least 75% of locations having a 1-σ range of less than 50 years for all epochs except the DACP. For this epoch, 33% of locations show a 1-σ range of over 100 years, mostly concentrated in a few Southern Hemisphere and tropical regions (Extended Data Fig. **2**).

Sensitivity tests for Figure 3

Additionally, we conducted a number of sensitivity tests. We tested using an alternative period length of 101-years (instead of 51 years) to calculate the running averages (Extended Data Fig. **5**) and using the full proxy network instead of the screened proxy network (Extended Data Fig. **4**). We computed epoch maps using the raw proxy data instead of the reconstructed fields (Extended Data Fig. **3**). We also ran reconstruction experiments using detrended calibration data to test for potential artifacts arising from the 20$^{\text{th}}$ century calibration period (Extended Data Fig. **8**).

There is also the potential concern that the proxy screening process and the reconstruction methodologies themselves could produce global coherence in the 20$^{\text{th}}$ century (Fig. **3**) given noisy proxies (thus no "real" underlying coherence). In this hypothetical scenario, signal-less proxies that by chance contain a trend in the 20$^{\text{th}}$ century could be selected in the screening process and weighted strongly in the reconstruction process, thus giving rise to a false sense of 20$^{\text{th}}$ century coherence.

To test this null hypothesis we generated three kinds of noise proxies that we then used within each of the reconstruction routines. For the first

kind, we generated noise proxies that are in the same locations and have the same AR spectrum and temporal coverage as the 210 real proxies used in our reconstructions[56,57]. This kind of noise proxy assesses the role that the reconstruction methodologies themselves have in potentially biasing the result of Fig. **3**. The second kind of proxies are the same as the first except that we additionally applied the R-FDR proxy-screening process to noise proxies representing the full PAGES2k database of 515 annually or higher resolved proxies[20]. This screening leaves on average about $n = 66$ noise proxies and may shed light on the influence of the screening approach on the results. As a final, even more conservative noise proxy experiment, we force-screened the noise proxies to have the same number ($n = 210$) as the screened real proxy data used herein. This means that we repeatedly generate noise-proxy time series at each location until the time series passes the R-FDR screening criterion. These second and third kind of proxies test the role that screening plays in the reconstruction process.

Of these three noise proxy experiments, we consider that the second ($n = 66$) is likely the most representative null experiment to compare the results of Fig. **3** against. While the third, $n = 210$ experiment uses the same number of input data, it does not accurately reflect the process of how proxy data are selected for the real data reconstructions. Repeatedly generating noise proxies at each location until a certain number of them pass the screening will increase the potential bias of the screening effect and will further amplify and build the observational signal into the proxy network, thereby eroding away the "noisiness" of the null.

We generated 25 sets of noise proxy networks for the three kinds of noise proxies and performed 100-member ensemble reconstructions for each reconstruction method, thus producing 45,000 global noise-proxy reconstructions. The results (Extended Data Fig. **7**) indicate that while it is possible in some proxy noise realizations to generate artificial 20$^{\text{th}}$ century warming coherence from the reconstruction algorithms themselves as well as from the proxy screening process, neither of these factors can explain the amount of 20$^{\text{th}}$ century warming coherence (panel **c**). We note also that the 20$^{\text{th}}$ century warming in the noise proxy reconstructions is a product of largely three reconstruction methods (GraphEM, DA, and AM, depending on the screening approach). We find that the noise-proxy reconstructions produce coherence that is always less than that seen in the real proxy reconstructions over all noise proxy experiments and methods. The median global area fraction showing largest ensemble probability for maximum 51-year temperatures within

the 20th century is between 37% and 67% for the noise proxies, depending on the screening approach, vs. 98% global area fraction for real proxies for both unscreened and screened networks (Extended Data Fig. **7c**) and climate model data (Extended Data Fig. **8**).

All these analyses, in conjunction with the independent verification using climate model data in Extended Data Fig. **8**, corroborate the results of Fig. **3** and show that they are robust to major technical choices.

Figure 4

While Fig. **3** is based on ensemble probabilities, Fig. **4** addresses the spatial consistency of warm and cool extremes within each ensemble member. To generate Fig. **4**, we first calculate 51-year running averages of each reconstruction ensemble member and identify the epoch peak-years (the maximum values for warm epochs or minimum values for cold epochs) at each grid point. Then we find which sliding 51-year period contains the most peak-years in terms of the global area fraction. The center year of this 51-year period is shown on the x-axis of Fig. **4**, while the y-axis is the area-weighted fraction of peak-years that are contained within the 51-year period. This process thus identifies the period when there is the largest spatial agreement about multi-decadal temperature extremes. For example, 24% of grid cells in the PCR ensemble member #1 have the LIA peak-year (Coldest 51-year average during the LIA epoch) within the period 1815-1865. No other 51-year period within the LIA epoch contains a higher area fraction of LIA minima in this ensemble member. The circle for this ensemble member is therefore drawn at the coordinates (1840,-0.24) in Fig. **4**. Boxplots on the right hand side of Fig. **4** integrate the area fractions of all ensemble members independent from the timing.

As a null-reference to test the significance of the area fractions, we repeated the Fig. **4** calculations based on multivariate random fields with realistic spatio-temporal covariance properties[15] using the `rmvn` function in the R package `mgcv`: First, the covariance matrix V of the ensemble median reconstruction field over the full 2,000 years is calculated for each reconstruction method. Second, a "square root" of V is generated using pivoted cholesky decomposition[58]. A matrix of normal white noise is then multiplied with the transpose of this square root matrix to obtain a multivariate normal matrix with the same covariance structure as the reconstructed field. Each grid-cell in this random field is then modified with an AR(1) model to obtain a time se-

ries with the same first order auto-regression coefficient as the corresponding grid cell in the ensemble median reconstruction[59]. This process is repeated 100 times for each reconstruction method to obtain a 600-member ensemble as in the real-world reconstructions. We then perform the search for the peak warming/cooling for each epoch in this noise-field ensemble using the same process a described above. The area fractions resulting from these noise-fields are shown with grey shading in Fig. **4**. We use the Mann-Whitney U test (α=0.05, one-tailed) to test if the area fractions in the reconstruction ensembles are significantly larger than expected from these noise fields (both n=600). Alternative benchmarks based on noise-proxy reconstructions are shown in Extended Data Fig. **6**. As above for Fig. **3**, all our sensitivity experiments (Extended Data Figs. **4**-**6**), confirm the robustness of our findings.

Methods References

31. Morice, C. P., Kennedy, J. J., Rayner, N. A. & Jones, P. D. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research* **117,** D08101 (2012).

32. Taylor, M. H., Losch, M., Wenzel, M. & Schröter, J. On the Sensitivity of Field Reconstruction and Prediction Using Empirical Orthogonal Functions Derived from Gappy Data. *Journal of Climate* **26,** 9194–9205 (2013).

33. Gneiting, T. & Raftery, A. E. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* **102,** 359–378 (2007).

34. Werner, J. P. & Tingley, M. P. Technical Note: Probabilistically constraining proxy age–depth models within a Bayesian hierarchical reconstruction model. *Clim. Past* **11,** 533–545 (2015).

35. Cook, E. R., Briffa, K. R. & Jones, P. D. Spatial regression methods in dendroclimatology: A review and comparison of two techniques. *International Journal of Climatology* **14,** 379–402 (1994).

36. Tipton, J., Hooten, M., Pederson, N., Tingley, M. & Bishop, D. Reconstruction of late Holocene climate based on tree growth and mechanistic hierarchical models. *Environmetrics* **27,** 42–54 (2016).

37. Werner, J. P., Divine, D. V., Charpentier Ljungqvist, F., Nilsen, T. & Francus, P. Spatio-temporal variability of Arctic summer temperatures over the past 2 millennia. *Clim. Past* **14,** 527–557 (2018).

38. Jones, P. *et al.* High-resolution palaeoclimatology of the last millennium: a review of current status and future prospects. *The Holocene* **19,** 3–49 (2009).

39. Mann, M. E. *et al.* Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proceedings of the National Academy of Sciences* **105,** 13252–13257 (2008).

40. Cook, E. R. *et al.* Asian Monsoon Failure and Megadrought During the Last Millennium. *Science* **328,** 486–489 (2010).

41. Neukom, R. *et al.* Inter-hemispheric temperature variability over the past millennium. *Nature Clim. Change* **4,** 362–367 (2014).

42. Luterbacher, J. *et al.* Reconstruction of sea level pressure fields over the Eastern North Atlantic and Europe back to 1500. *Climate Dynamics* **18,** 545–561 (2002).

43. Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M. & Wanner, H. European seasonal and annual temperature variability, trends, and extremes since 1500. *Science* **303,** 1499–1503 (2004).

44. Neukom, R. *et al.* Multi-centennial summer and winter precipitation variability in southern South America. *Geophysical Research Letters* **37,** L14708 (2010).

45. Neukom, R. *et al.* Multiproxy summer and winter surface air temperature field reconstructions for southern South America covering the past centuries. *Climate Dynamics* **37,** 35–51 (2011).

46. Smerdon, J. E. & Pollack, H. N. Reconstructing Earth's surface temperature over the past 2000 years: the science behind the headlines. *Wiley Interdisciplinary Reviews: Climate Change* **7,** 746–771 (2016).

47. Christiansen, B., Schmith, T. & Thejll, P. A Surrogate Ensemble Study of Climate Reconstruction Methods: Stochasticity and Robustness. *Journal of Climate* **22,** 951–976 (2009).

48. Smerdon, J. E., Kaplan, A., Chang, D. & Evans, M. N. A Pseudoproxy Evaluation of the CCA and RegEM Methods for Reconstructing Climate Fields of the Last Millennium. *Journal of Climate* **23,** 4856–4880 (2010).

49. Guillot, D., Rajaratnam, B. & Emile-Geay, J. Statistical paleoclimate reconstructions via Markov random fields. *The Annals of Applied Statistics* **9,** 324–352 (2015).

50. Friedman, J., Hastie, T. & Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9,** 432–441 (2008).

51. Steiger, N. J., Hakim, G. J., Steig, E. J., Battisti, D. S. & Roe, G. H. Assimilation of Time-Averaged Pseudoproxies for Climate Reconstruction. *Journal of Climate* **27,** 426–441 (2014).

52. Hakim, G. J. *et al.* The last millennium climate reanalysis project: Framework and first results. *Journal of Geophysical Research: Atmospheres* **121,** 6745–6764 (2016).

53. Steiger, N. J., Smerdon, J. E., Cook, E. R. & Cook, B. I. A reconstruction of global hydroclimate and dynamical variables over the Common Era. *Scientific Data* **5,** 180086 (2018).

54. Otto-Bliesner, B. L. *et al.* Climate Variability and Change since 850 CE: An Ensemble Approach with the Community Earth System Model. *Bulletin of the American Meteorological Society* **97,** 735–754 (2016).

55. Gómez-Navarro, J. J., Zorita, E., Raible, C. C. & Neukom, R. Pseudoproxy tests of the analogue method to reconstruct spatially resolved global temperature during the Common Era. *Clim. Past* **13,** 629–648 (2017).

56. Hosking, J. R. M. Modeling persistence in hydrological time series using fractional differencing. *Water Resources Research* **20,** 1898–1908 (1984).

57. Wahl, E. R. & Smerdon, J. E. Comparative performance of paleoclimate field and index reconstructions derived from climate proxies and noise-only predictors. *Geophysical Research Letters* **39** (2012).

58. Higham, N. J. Cholesky factorization. *Wiley Interdisciplinary Reviews: Computational Statistics* **1,** 251–254 (2009).

59. Neukom, R., Schurer, A. P., Steiger, N. J. & Hegerl, G. C. Possible causes of data model discrepancy in the temperature history of the last Millennium. *Scientific Reports* **8,** 7572 (2018).

Data availability

The PAGES 2k, v2.0.0 dataset is archived at the World Data Service for Paleoclimatology (hosted by NOAA) formatted for both LiPD and WDS ASCII template (https://www.ncdc.noaa.gov/paleo/study/21171). The screened input data matrix, instrumental target grid, and the reconstruction outcomes from this study are available at Figshare (doi:10.6084/m9.figshare.c.4498373.v1) and NOAA-WDS Paleoclimatology (www.ncdc.noaa.gov/paleo). We strongly recommend using the multi-method ensembles, when working with the reconstructions. For analyses of global mean temperatures we recommend using the reconstruction of the PAGES 2k companion project explicitly targeting the global mean (PAGES2k Consortium, in press).

Code availability

The code to generate the figures is available with the output data (see above).

Extended Data Legends

Extended Data Figure 1 | Sensitivity plots for Fig. 2. Coloured areas show global area fractions with warm (red shading) and cold (blue shading) temperature anomalies with respect to a 1-2000 CE reference period (see Methods). **a,b** Same as Fig. **2** (annual vs. 51-year filtered data) but over the full ensemble. While Fig. **2** shows the mean area fractions of the six reconstruction ensemble medians, this Figure displays the percentage over all ensemble members and locations. Similar to the weak spatial coherence in Figs. **3** and **4**, this ensemble based illustration shows much weaker spatial coherence (lower percentages) than Fig. **2**. **c** Same as Fig. **2b** but for 101-year filtered data instead of 51-year filterd data.

Extended Data Figure 2 | Uncertainties in epoch timings. a-e Uncertainty of the century of peak warming/cooling (Fig. **3**) at each location and for each epoch is quantified by bootstrapping. The maps display the standard deviation of the 1,000 re-calculations of the century with largest ensemble probability for peak 51-year warming/cooling based on bootstrap-resampling of the 600 ensemble members (Methods). They show that the identified cold and warm peaks are generally robust across epochs. The largest uncertainties are found for the DACP epoch and tropical and SH regions. Also, the uncertainty for the CWP warming is very large in the (mainly Antarctic) regions, where peak warming is not identified in the late 20^{th} century (Fig. 3). **f-j** same as **a-e** but showing the 90% range instead of the standard deviation.

Extended Data Figure 3 | Epoch timings in proxy data. Peak warming/cooling for each epoch in the proxy records. **a-e** All proxies with full coverage of the respective epoch are shown. **f-i** Same as **a-e** but showing also proxies with only partial coverage of the respective epoch. In contrast to Fig. **3**, colors are coded by century using a differential color scheme for better visibility. Relative fraction of proxy records with peak warming/cooling in each century is indicated with barplots above the legend strips. Note that for this figure, we use the full, unscreened PAGES2 k v2.0.0 proxy database (the screened network yields a consistent picture). This analysis shows that the heterogeneity in the timing of maxima and minima is an inherent property of the input proxy data itself, which show a similar lack of global coherence in the timing of each putative pre-industrial climate epoch. Though note that the proxy maps are not directly comparable to the reconstruction maps because the reconstructions are an objectively weighted, statistically optimal fit between all available proxy values using covariance information from a spatial temperature field.

Extended Data Figure 4 | Un-screened proxy network. a-e Same as Fig. **3** but using the full unscreened PAGES 2k temperature proxy database. Note that the methods used herein do not incorporate low-frequency records (resolution >1 year), therefore only 559 of the 692 records from the PAGES 2k database are used to generate this figure. Colours in maps indicate the century with the largest ensemble probability of containing the warmest (**a,b,c**) and coldest (**d,e**) 51-year period within each climatic epoch (see Methods). **f** Same as Fig. **4** but using the full unscreened PAGES 2k temperature proxy database. Global area fraction (y-axis) that simultaneously experienced the warmest (top) or coldest (bottom) multi-decadal period (51 years) during each of five different epochs (see Methods). Each solid circle is an ensemble member plotted according to the year (x-axis) in which the largest area experienced peak warm/cold conditions. Gray shading represents the distribution from the same analysis based on multivariate AR1 noise fields, with darker colors indicating higher probability. Boxplots on the right show area fractions integrated over time. The center line is the median, the ends of the boxes represent the interquartile range, and whiskers the 90% range. Bold text represents epochs with reconstructed area fractions significantly higher than from the noise fields (Mann-Whitney U test, α=0.05). Recall that CWP maxima are searched within the full 2,000 year reconstruction period. In contrast to using the screened proxy matrix displayed in Fig. **4**, the period of largest warming within this 2,000 year range falls in the 2nd century CE for one single CCA ensemble member, thus overlapping with the search windows of the RWP period. Therefore, circles representing the CWP have a black border to allow distinction from other epochs.

Extended Data Figure 5 | 101-year maxima and minima. **a-e** Same as Fig. **3** but for 101-year instead of 51-year periods. Colours in maps indicate the century with the largest ensemble probability of containing the warmest (**a,b,c**) and coldest (**d,e**) 51-year period within each climatic epoch (see Methods). **f** Same as Fig. **4** but for 101-year periods. Global area fraction (y-axis) that simultaneously experienced the warmest (top) or coldest (bottom) multi-decadal period (51 years) during each of five different epochs (see Methods). Each solid circle is an ensemble member plotted according to the year (x-axis) in which the largest area experienced peak warm/cold conditions. Gray shading represents the distribution from the same analysis based on multivariate AR1 noise fields, with darker colors indicating higher probability. Boxplots on the right show area fractions integrated over time. The center line is the median, the ends of the boxes represent the interquartile range, and whiskers the 90% range. Bold text represents epochs with reconstructed area fractions significantly higher than from the noise fields (Mann-Whitney U test, α=0.05). Recall that CWP maxima are searched within the full 2,000 year reconstruction period. In contrast to the 51-year maxima displayed in Fig. **4**, some of the 101-year maxima within this 2,000 year range fall within the pre-1350 period, thus overlapping with the search windows of the RWP and MCA periods. Therefore, circles representing the CWP have a black border to allow distinction from other epochs.

Extended Data Figure 6 | Alternative benchmarks for the spatial consistency of epoch timings. Same as Fig. **4** but overlaid with results from PCR AR-noise-proxy reconstructions (see also Extended Data Fig. **7**) instead of gray bars based on AR(1)-noise fields. Global area fraction (y-axis) that simultaneously experienced the warmest (top) or coldest (bottom) multi-decadal period (51 years) during each of five different epochs (see Methods). Each solid circle is an ensemble member plotted according to the year (x-axis) in which the largest area experienced peak warm/cold conditions. The result from each noise-proxy reconstruction ensemble member is shown as a gray circle. Noise proxy reconstruction circles for the CWP epoch have a black border to allow distinction with the RWP and MCA epochs because the AR-noise proxy results are scattered through time. Boxplots on the right hand side integrate the area fractions of all ensemble members independent from the timing. The center line is the median, the ends of the boxes represent the interquartile range, and whiskers the 90% range. Note that the area fractions for the noise proxy reconstructions are lower than for the AR-noise fields in Fig. **4**, but still only the CWP epoch stands out as having significantly larger fractions than the noise benchmark. Dotted horizontal black lines indicate the area fractions expected from within a spatio-temporally uncorrelated field. In this case, the expected area fraction is modeled with a binomial distribution with $M = 2592$ trials (the number of grid cells) and the probability of success on each trial being $51/N$, where N is the number of years, within which the 51-year peak is searched for each epoch. The dotted lines represent the 95^{th} percentile of this distribution divided by M.

Extended Data Figure 7 | Timing of peak warming in noise-proxy reconstructions. a Same as the CWP-panel in Fig. **3** but for reconstructions using noise proxies. Colours in maps indicate the century with the largest ensemble probability of containing the warmest 51-year period within the CE (see Methods). Maps show the 25 reconstruction realizations, each consisting of 6×100-member ensemble reconstructions, for the R-FDR screened ($n = 66$) noise proxy networks (see Methods). **b** The global area fraction of peak warmth in each century for each reconstruction method. Top: real proxies (screened). Middle: average values across the 25 screened ($n = 66$) noise proxy reconstruction ensembles. bottom: average values across the 25 force-screened ($n = 210$) noise proxy reconstruction ensembles. **c** Fraction of global area having the CWP warm peak within the 20$^{\text{th}}$ century for all three noise proxy types described in the Methods. Large gray boxplots: noise proxy reconstructions across all methods. Gray filled circles: individual noise proxy reconstructions across all methods. Coloured boxplots: Noise proxy results for the individual reconstruction methods (colours as in **b**). Vertical red lines: real proxy reconstructions for both unscreened (upper panel) and screened (lower two panels) networks. Boxplots are across 25 reconstruction experiments; center line = median, boxes = interquartile range, whiskers = 95% range. All noise proxy experiments across all methods yield a weaker spatial agreement of maximum 20$^{\text{th}}$ centruy 51-year warming compared to the real data reconstructions. The more "traditional" statistical reconstruction methods (CPS, PCR and CCA) mostly exhibit smaller areas of 20$^{\text{th}}$ century warming in the noise reconstructions compared to the other methods (GraphEM, AM and DA). A possible explanation for this difference is the fact that the traditional group of methods is designed to yield reconstructions with as little variance loss as possible independent from data uncertainty (see e.g., Ref. 60). Reconstructed temperatures over the full CE thus exhibit fluctuations with a magnitude comparable to the calibration period in all noise experiments. In contrast, the newer methods usually generate reconstructed variance that is inversely proportional to the errors in the input data. Thus they converge towards zero with increasing data uncertainty and decreasing coherence among the input data, as is the case in the noise proxy experiments. For these methods, this results in noise proxy reconstructions with little temperature variance prior to the calibration period and thus a higher probability that the 20$^{\text{th}}$ century warming exceeds earlier warm periods in magnitude. For a general discussion of the results see Methods section.

Extended Data Figure 8 | Timing of peak warm and cold periods in detrended calibration reconstructions and model data. Colours in maps indicate the century with the largest ensemble probability of containing the warmest (**a,b,c,e**) and coldest (**d**) 51-year period within each climatic epoch (see Methods). **a** Same as the CWP-panel in Fig. **3** but including barplots above the legend strip showing the relative occurrence of peak warming in each century for each reconstruction method. **b** Same as **a** but using only the CPS reconstruction with calibration based on linearly detrended proxy and instrumental target data. Detrending calibration data partly removes variance associated with physical processes[61] leading to reduced reconstruction skill (see Refs 62-63 for detailed discussion of the topic). Nevertheless, the reconstruction based on detrended data shows warm peaks in the 20th century over much of the globe with the exception of the Eurasian land masses and the Southern Hemisphere extratropics. **c-e** same as Fig. **3 a, e, c** but using climate model simulations. We use CMIP5[66] last millennium runs from the models BCC-CSM[67], CCSM4[68], CESM-LME[54] (member 10), CSIROMk3L-1-2[69], GISS-E2-R[70] (member 127), HadCM3[71], IPSL-CM5A-LR[72] and MPI_ESM_P[73]. From models with more than one ensemble member, only one member is used, to avoid biases towards single models. Note that because the shortest simulations extend back to 851, no results are available for the RWP and DACP periods, and the MCA peak is only searched within the period 851 to 1350.

Extended Data Figure 9 | Reconstruction skill. a Validation metrics (see Methods) for the different reconstruction methods. Boxplots integrate over all grid cells, the center line is the median, the ends of the boxes represent the interquartile range, and whiskers the 90% range. Horizontal axes are adjusted so that better skill is always on the right hand side. Dotted vertical lines represent the median across all grid-cells and methods. Dashed gray indicates the value of zero (except RMSE). Note that over the 1881-1910 validation period, the spatial coverage of instrumental data is already very sparse[31,74], strongly limiting the validity of verification experiments at the grid cell level. Also, the limited number of years available for validation can strongly affect the outcome of skill estimates[75–78]. Note that the short validation period does not allow for a robust assessment of reconstruction skill on decadal and lower frequencies. The validation statistics are representative of the most replicated proxy nest. Extending these estimates back in time requires a nested reconstruction approach, which is not implemented in all methods used herein. Extended Data Fig. **10**, shows the corresponding values for the years 1 and 1000 CE for the PCR method. The width of the uncertainty intervals shown in Fig. **1** (red line) provides an illustration of the continuously increasing reconstruction errors back in time. **b,c,d,e** Maps showing the spatial distribution of skill scores. The mean of all methods is shown. Darker red refers to higher skill in all maps. Proxy locations are indicated with gray circles. In general, the reconstruction skill is lowest in the high southern latitudes, tropical South America and Africa and over some oceanic regions where proxy data coverage, but also instrumental data availability, is sparse.

Extended Data Figure 10 | Reconstruction skill and methods agreement in years 1 CE and 1000 CE. a Density of CRPS_RE values in PCR reconstructions based on the full proxy network (dark yellow, same as the PCR boxplot in Extended Data Fig. **9**), the proxy records extending at least to 1000 CE (green) and the records covering the full CE (blue). Numbers besides the curves indicate the percentage of grid cells with positive values. **b** (**c**) maps showing the spatial distribution of CRPS_RE for the years 1000 CE (1 CE). Proxy locations are indicated with gray circles. **d,e,f** Same as **a,b,c** but for the CRPS_CE. **g,h,i** Same as **a,b,c** but for the RMSE. **j,k,l** Same as **a,b,c** but for the correlation coefficient. In general, the spatial patterns remain similar over time but with areas of lower skill naturally extending back in time (see also the red line in Fig.**1**). The largest decrease in skill generally occurs in the first Millennium CE. **m,n,o** Average correlation of ensemble median reconstructions accross all methods over the period 1900-1999, 1000-1099 and 1-99, respectively. More than 99% of correlations are positive in all three periods. 97%, 76% and 73% of correlations in the 20^{th}, 11^{th} and 1^{st} century CE, respectively, are above 0.28, which is the average $\alpha = 0.05$ significance level given the autocorrelation in the reconstructions. In all periods the method agreement is larger in the Northern Hemisphere, particularly in the North Pacific and European domains. Lowest agreement is found over tropical South America and Africa and over the Southern Ocean, the same areas that also exhibit the largest errors in the reconstructions.

Extended Data References

60. Christiansen, B. & Ljungqvist, F. C. Challenges and perspectives for large-scale temperature reconstructions of the past two millennia. *Reviews of Geophysics* **55,** 40–96 (2017).

61. Ammann, C. M. & Wahl, E. R. The importance of the geophysical context in statistical evaluations of climate reconstruction procedures. *Climatic Change* **85,** 71–88 (2007).

62. Gergis, J., Neukom, R., Gallant, A. J. E. & Karoly, D. J. Australasian Temperature Reconstructions Spanning the Last Millennium. *Journal of Climate* **29,** 5365–5392 (2016).

63. Wahl, E. R., Ritson, D. M. & Ammann, C. M. Comment on "Reconstructing Past Climate from Noisy Data". *Science* **312,** 529–529 (2006).

64. Von Storch, H. Reconstructing Past Climate from Noisy Data. *Science* **306,** 679–682 (2004).

65. Bürger, G. & Cubasch, U. Are multiproxy climate reconstructions robust? *Geophysical Research Letters* **32** (2005).

66. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society* **93,** 485–498 (2012).

67. Xiao-Ge, X., Tong-Wen, W. & Jie, Z. Introduction of CMIP5 Experiments Carried out with the Climate System Models of Beijing Climate Center. *Advances in Climate Change Research* **4,** 41–49 (2013).

68. Landrum, L. *et al.* Last Millennium Climate and Its Variability in CCSM4. *Journal of Climate* **26,** 1085–1111 (2013).

69. Phipps, S. J. *et al.* The CSIRO Mk3L climate system model version 1.0 – Part 2: Response to external forcings. *Geosci. Model Dev.* **5,** 649–682 (2012).

70. Schmidt, G. A. *et al.* Present-day atmospheric simulations using GISS ModelE: Comparison to in situ, satellite, and reanalysis data. *Journal of Climate* **19,** 153–192 (2006).

71. Schurer, A. P., Hegerl, G. C., Mann, M. E., Tett, S. F. B. & Phipps, S. J. Separating Forced from Chaotic Climate Variability over the Past Millennium. *Journal of Climate* **26,** 6954–6973 (2013).

72. Dufresne, J.-L. *et al.* Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Climate Dynamics* **40,** 2123–2165 (2013).

73. Jungclaus, J. H. *et al.* Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model. *Journal of Advances in Modeling Earth Systems* **5,** 422–446 (2013).

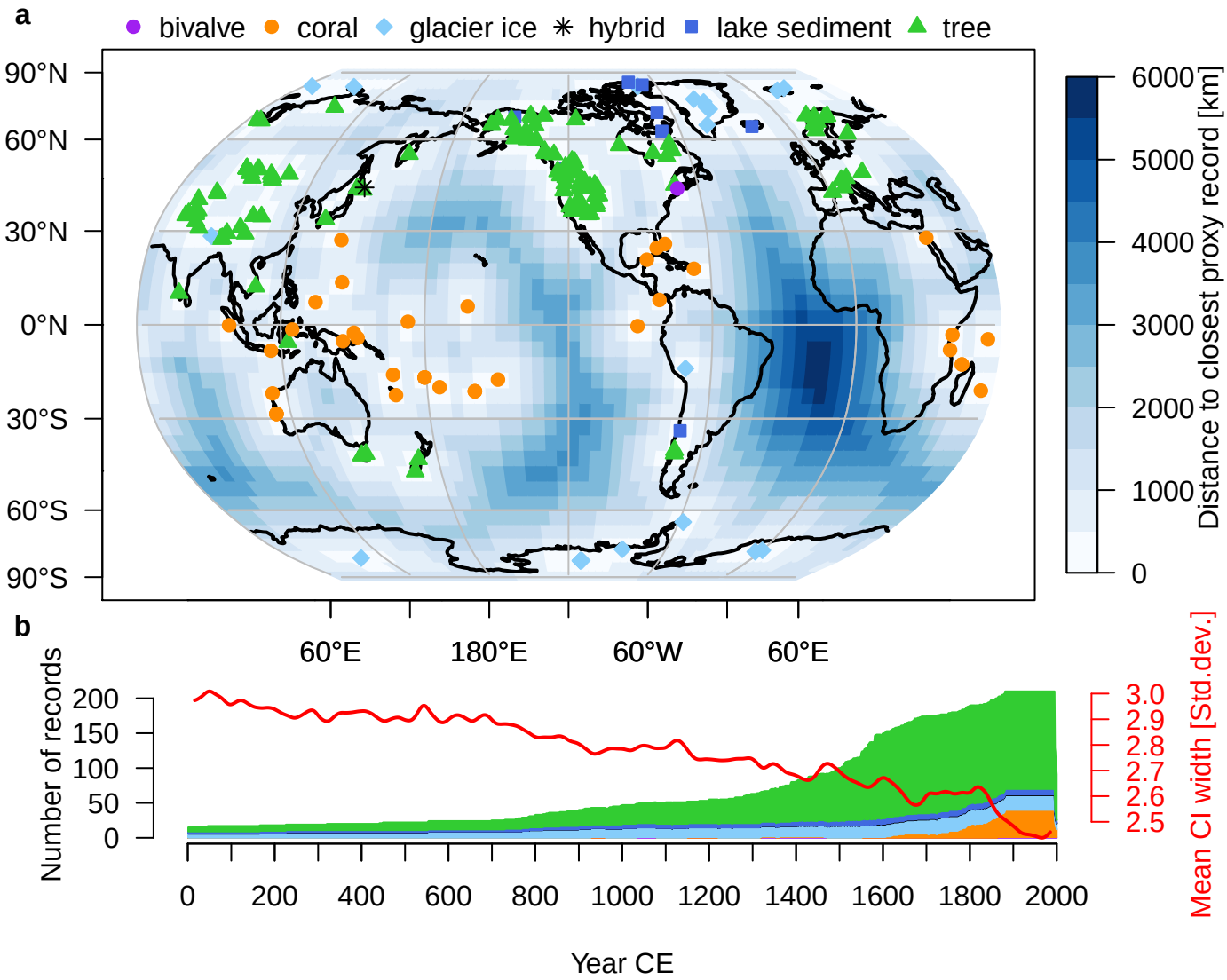
74. Cowtan, K. & Way, R. G. Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society* **140,** 1935–1944 (2014).

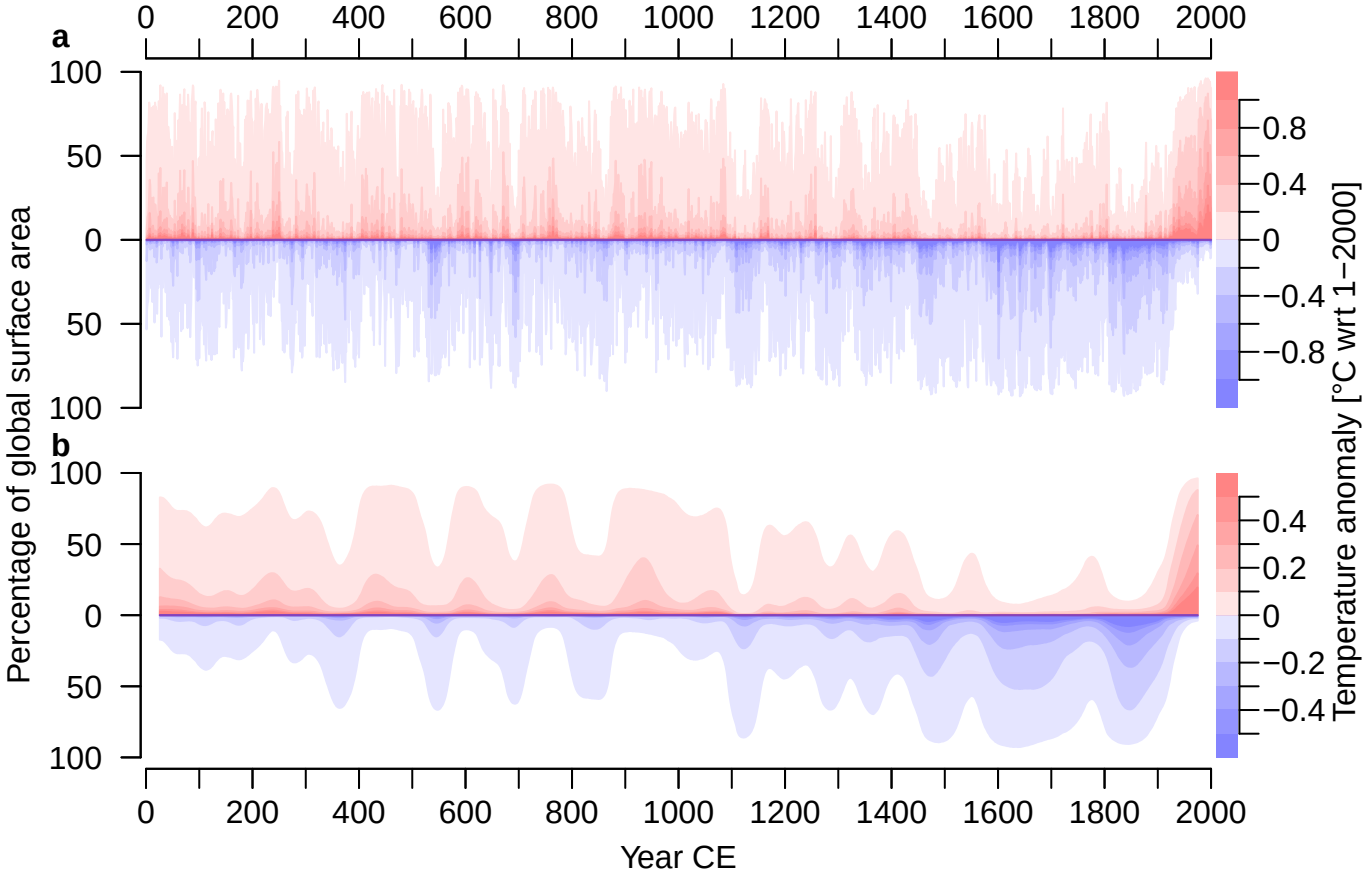
75. Mann, M. E., Rutherford, S., Wahl, E. & Ammann, C. Robustness of proxy-based climate field reconstruction methods. *Journal of Geophysical Research* **112,** D12109 (2007).

76. Gallant, A. J. E. & Gergis, J. An experimental streamflow reconstruction for the River Murray, Australia, 1783-1988. *Water Resources Research* **47,** W00G04 (2011).

77. Gergis, J. *et al.* On the long-term context of the 1997–2009 'Big Dry' in South-Eastern Australia: insights from a 206-year multi-proxy rainfall reconstruction. *Climatic Change* **111,** 923–944 (2012).

78. Frank, D. C. *et al.* Ensemble reconstruction constraints on the global carbon cycle sensitivity to climate. *Nature* **463,** 527–530 (2010).


a
bivalve
coral
glacier ice
hybrid
lake sediment
tree
90°N
60°N
30°N
0°N
30°S
60°S
90°S
60°E
180°E
60°W
60°E
6000
5000
4000
3000
2000
1000
0
Distance to closest proxy record [km]
b
Number of records
200
150
100
50
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Year CE
3.0
2.9
2.8
2.7
2.6
2.5
Mean CI width [Std.dev.]


a
b
0
200
400
600
800
1000
1200
1400
1600
1800
2000
100
50
0
50
100
Percentage of global surface area
Year CE
0.8
0.4
0
−0.4
−0.8
0.4
0.2
0
−0.2
−0.4
Temperature anomaly [°C wrt 1−2000]

Roman Warm Period [1–750]

a

60 180 −60 60

90 60 30 0 −30 −60 −90

0 200 400 600 800

Year CE

Medieval Climate Anomaly [751–1350]

b

60 180 −60 60

700 800 900 1100 1300

Year CE

Current Warm Period [1–2000]

c

60 180 −60 60

90 60 30 0 −30 −60 −90

0 500 1000 1500 2000

Year CE

Dark Ages Cold Period [1–1000]

d

60 180 −60 60

90 60 30 0 −30 −60 −90

0 200 400 600 800 1000

Year CE

Little Ice Age [1001–2000]

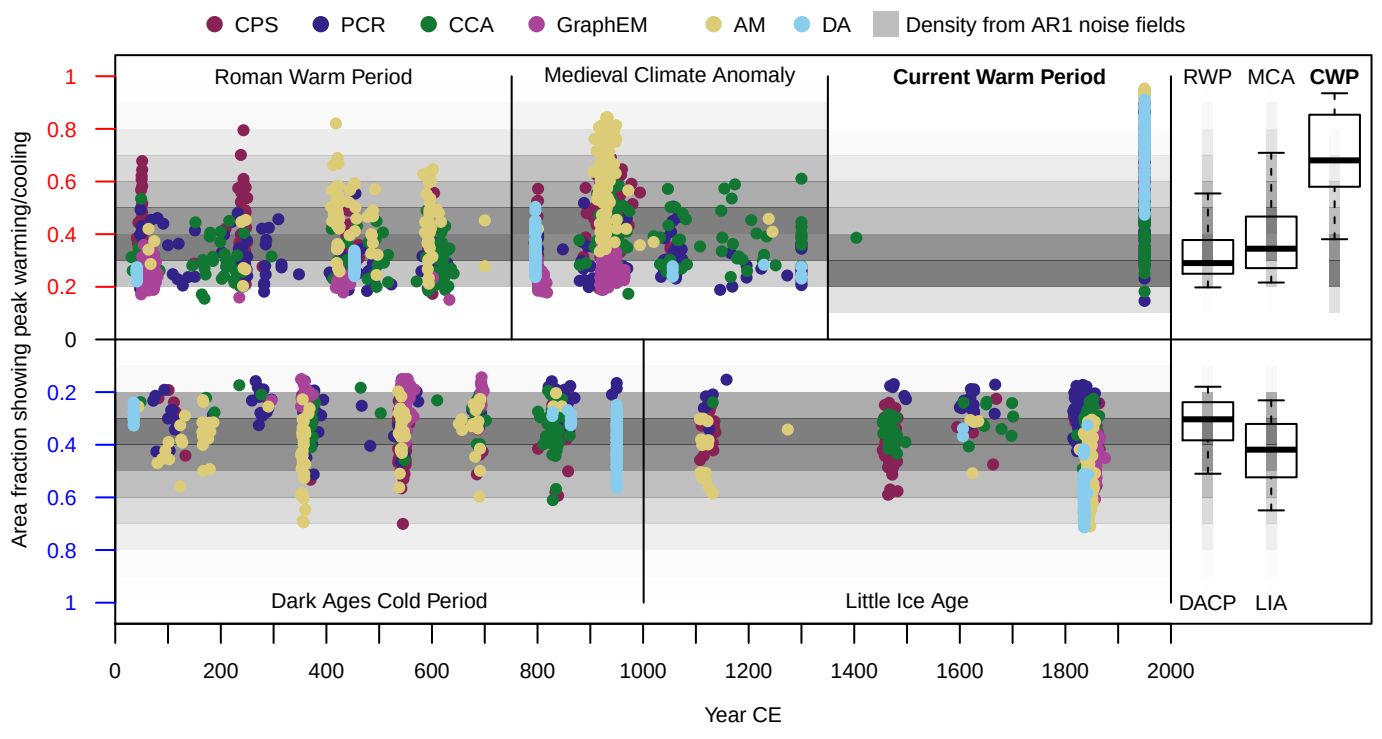
e

60 180 −60 60

90 60 30 0 −30 −60 −90

1000 1200 1400 1600 1800 2000

Year CE


CPS
PCR
CCA
GraphEM
AM
DA
Density from AR1 noise fields
Roman Warm Period
Medieval Climate Anomaly
Current Warm Period
RWP
MCA
CWP
Dark Ages Cold Period
Little Ice Age
DACP
LIA
Area fraction showing peak warming/cooling
Year CE
0
200
400
600
800
1000
1200
1400
1600
1800
2000
1
0.8
0.6
0.4
0.2
0