

Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era

PAGES 2k Consortium*

Multidecadal surface temperature changes may be forced by natural as well as anthropogenic factors, or arise unforced from the climate system. Distinguishing these factors is essential for estimating sensitivity to multiple climatic forcings and the amplitude of the unforced variability. Here we present 2,000-year-long global mean temperature reconstructions using seven different statistical methods that draw from a global collection of temperature-sensitive palaeoclimate records. Our reconstructions display synchronous multidecadal temperature fluctuations that are coherent with one another and with fully forced millennial model simulations from the Coupled Model Intercomparison Project Phase 5 across the Common Era. A substantial portion of pre-industrial (1300–1800 CE) variability at multidecadal timescales is attributed to volcanic aerosol forcing. Reconstructions and simulations qualitatively agree on the amplitude of the unforced global mean multidecadal temperature variability, thereby increasing confidence in future projections of climate change on these timescales. The largest warming trends at timescales of 20 years and longer occur during the second half of the twentieth century, highlighting the unusual character of the warming in recent decades.

Arising from both natural forcing and unforced internal variability in the climate system, decadal- to multidecadal surface temperature variations will continue to be superimposed on the anthropogenic global warming trend^{1–7}, affecting estimates of transient climate response to individual forcings and regional climate change⁸. Knowledge of multidecadal global mean surface temperature (GMST) fluctuations, as well as their relative magnitudes compared with that of the projected long-term warming trend, is thus of crucial importance for understanding future climate change^{9,10}.

Global and regional-scale multidecadal temperature variability (MDV) was found to be under- or overestimated in climate simulations of the recent past^{11–13}. Some multicentennial to millennial-scale assessments of palaeoclimate observations and simulations suggest that temperature variability at multidecadal and lower frequencies is underestimated by the Coupled Model Intercomparison Project Phase 5 (CMIP5)¹⁴ models^{15,16}, contrasting with other evidence suggesting that models contain realistic amplitudes of MDV^{17,18}. Global networks of instrumental observations extend back to 1850 at best¹⁹, which is too short a period to fully characterize MDV and its driving forces and evaluate the simulated MDV amplitude²⁰. Estimating the relative contributions of various climate forcings, and assessing the ability of climate models to accurately simulate observed climate phenomena, requires a far longer perspective¹⁰.

Over the Common Era (CE, the past 2,000 yr), palaeoclimate proxy-based observations of temperature and climate forcings are available at up to sub-annual resolution and cover much of the globe²¹. Reconstructed GMST over the full Common Era provides context for recent warming, and offers an opportunity to benchmark the ability of climate models to simulate MDV and to study the planetary energy balance and other key aspects of the climate system. Most previous large-scale millennium-length temperature reconstructions have highlighted the extraordinary rate and magnitude of the warming in recent decades¹. However, there are

considerable discrepancies among the reconstructions in the amplitude, and partly the timing, of past temperature fluctuations at inter-annual to multicentennial timescales^{1,22}. The extent to which these differences can be attributed to different reconstruction methodologies relative to different palaeoclimate proxy networks is unclear, although both have been shown to strongly influence temperature reconstructions²³. The recently updated compilation of temperature-sensitive records in a community-vetted, quality-controlled database²¹ provides an opportunity to improve the reconstruction of GMST for the Common Era. Despite much innovative work in this area^{22,24}, there have been few coordinated efforts to reconstruct global temperatures using a consistent experimental framework²³.

Here, we use seven different statistical methods to reconstruct GMST over the past 2,000 yr (1–2000 CE; Table 1). The methods range from basic composite-plus-scaling (CPS) and regression-based techniques (such as principal component regression (PCR) and regularized errors in variables (M08)) frequently used in past reconstructions, to newer linear methods (optimal information extraction (OIE) and a Bayesian hierarchical model (BHM), for example) and techniques that account for nonlinear relations between proxy values and temperature (pairwise comparison; PAI) or combine information from proxy data and climate models (offline data assimilation; DA). All procedures use the same input dataset (ref. ²¹, Supplementary Fig. 1) and the same calibration dataset for the reconstruction target, the infilled version²⁵ of HadCRUT4¹⁹. For each method, a 1,000-member ensemble of reconstructions is generated, allowing a probabilistic assessment of uncertainties. Further details on each method, such as the various approaches used to generate the reconstruction ensembles and the ability to include low-resolution records, are provided in the Methods.

Temperature history of the Common Era

The temperature evolutions reconstructed by the seven methods (Fig. 1a) exhibit the major features of previous reconstructions

*A full list of authors and their affiliations appears at the end of the paper.

Table 1 | Overview of the reconstruction methods and data-model comparison

Method	Reference(s)	Includes only annually resolved proxies	Model/data variance ratios	Model versus data correlations
CPS	56,66	Yes	0.96 [0.39, 2.49]	0.64 [0.44, 0.80] (99.6%)
PCR	56,67	Yes	1.01 [0.39, 2.65]	0.60 [0.38, 0.77] (99.4%)
OIE	68	No	1.13 [0.48, 3.02]	0.61 [0.43, 0.83] (99.5%)
M08	66	No	1.01 [0.41, 2.59]	0.65 [0.49, 0.79] (99.8%)
PAI	69	No	0.63 [0.25, 1.65]	0.63 [0.43, 0.79] (99.7%)
BHM	70	Yes	1.12 [0.47, 2.91]	0.53 [0.37, 0.66] (98.9%)
DA	71	Yes	1.15 [0.44, 3.02]	0.62 [0.40, 0.86] (99.6%)

Values are ensemble medians and 2.5th and 97.5th percentiles are shown in square brackets. Percentage of ensemble members with correlations larger than expected from AR(1) noise in parenthesis. Unfiltered data and results for each individual model simulation are shown in Supplementary Tables 1–3.

over the Common Era¹. All reconstructions show that the GMST during the first millennium of the Common Era was warmer than during the second millennium (excluding the twentieth century). All reconstructions show a significant cooling trend before 1850 CE (Methods), followed by rapid industrial-era warming. The warmest 10 yr (30 yr, 50 yr) period of the past two millennia falls within the second half of the twentieth century in 94% (89%, 84%) of ensemble members. The most pronounced difference between the reconstructions produced by the different methods is the amplitude of the pre-industrial multicentennial cooling trend, manifesting in a temperature difference of about 0.5 °C between the warmest (DA method) and the coldest (BHM method) estimates around 1600 CE.

Methods that are implemented with proxy records of lower-than-annual resolution (Table 1 and Methods) yield systematically larger pre-industrial cooling trends, on average $-0.23^{\circ}\text{Ckyr}^{-1}$ (-0.31 , -0.11) (the ensemble median (2.5th percentile, 97.5th percentile)), compared with the other methods ($-0.09^{\circ}\text{Ckyr}^{-1}$ (-0.27 , 0.02)).

The uncertainties for all reconstruction methods increase backwards in time and are particularly large before medieval times (Supplementary Fig. 3), owing to the decreasing numbers of input proxy data (Supplementary Fig. 1) and associated changes in the observing network. The magnitude of uncertainties varies among the reconstructions depending on the method-specific approach to generating the reconstruction ensemble (Methods).

In the currently available database, many palaeoclimate archives are seasonally biased^{22,26,27} (Supplementary Section 3) and the relative seasonal representation varies over time. The majority of the included tree-ring records are detrended for removal of biological age effects by methods that cannot capture centennial- to multicentennial-scale temperature variability^{28,29}. This problem probably manifests more backwards in time²², and may result in an underestimation of low-frequency variability—especially during the first millennium of the Common Era³⁰. On the other hand, low-resolution marine records seem to overestimate the true variance³¹. Many of the palaeoclimate records also have a mixed and, presumably, nonlinear response to temperature and hydroclimate^{32,33} that is probably not stable over time³⁴.

Although multicentennial- to millennial-scale trends differ among the GMST reconstruction methods^{22,35,36} (Fig. 1a), multidecadal fluctuations are more similar (Fig. 1b). Removing the centennial-scale trend reveals remarkably coherent MDV with narrower confidence ranges across reconstruction techniques (Fig. 1b), suggesting a more robust result across methods.

Multidecadal variability and response to forcing

To explore the influence of the major known climate forcings on multidecadal GMST, we compared an ensemble of 23 CMIP5 climate model simulations (Methods) with each of the temperature reconstructions. We find strong reconstruction–model agreement

over the past millennium (Fig. 2) in both the timing and magnitude of multidecadal (30–200 yr bandpass-filtered) GMST variance: the median model/data variance ratio across all reconstruction and model ensemble members is 1.01 and close to one for the individual reconstruction methods (Table 1). Multidecadal data–model coherence is substantially larger than expected by chance (Table 1, Methods).

Warm anomalies around 1320, 1420, 1560 and 1780 CE, cold anomalies around 1260, 1450 and 1820 CE and the periods of reduced variability during the overall cold seventeenth century and the relatively warm eleventh century are captured by both the reconstructions and simulations. These similarities between reconstructions and models suggest a dominant influence of external forcing on multidecadal GMST variability.

The data–model agreement is particularly strong between 1300 and 1800 CE. The weaker agreement before 1300 CE might be mostly explained by the reduced quality of volcanic forcing estimates used in CMIP5 during this period³⁷; for example the 1109 CE eruption, which is followed by a clear GMST cooling (Fig. 2), is missing in the forcing datasets^{38,39} used for these CMIP5 simulations (Supplementary Fig. 20b). The apparent data–model mismatch in the nineteenth century is caused by the difference in response to the eruptions in 1809 and 1815 CE, which is probably overestimated by the CMIP5 model simulations⁴⁰ (Supplementary Fig. 19).

To quantitatively evaluate the GMST response to external forcings in the reconstructions compared with model simulations, we apply a formal detection and attribution (D&A) analysis^{41–43}. The D&A analysis uses robust linear regression of reconstructed GMST on unforced, single and cumulatively forced simulated GMST. The regression coefficients (scaling factors) may be interpreted as the extent to which reconstructed variability is explained by candidate unforced and forced responses. We choose 1300–1800 CE for the D&A analysis, as during this period the anthropogenic forcing is negligible, and the uncertainties in the forcings, simulations and reconstructions are relatively small and time-invariant compared with earlier centuries^{21,37,43} (Fig. 2).

For all reconstruction methods, the D&A scaling factors are significantly above zero when performing the D&A analysis on multidecadal timescales and using fully forced model simulations (Fig. 3a, left; for details see the Methods). This means that the influence of external forcing on GMST variance is detectable in the simulations. The full ensemble median of the scaling factors (0.89) is close to unity and in four of the seven reconstruction methods (CPS, PCR, PAI and BHM), the 90% range of scaling factors does encompass unity, indicating a consistent response to total external forcing in models and reconstructions. This result is robust for a single model (CESM-CAM5⁴⁴) and multimodel ensembles (Supplementary Fig. 7).

To disentangle the influence of different forcing factors, we repeat the D&A analysis by simultaneously regressing the ensemble

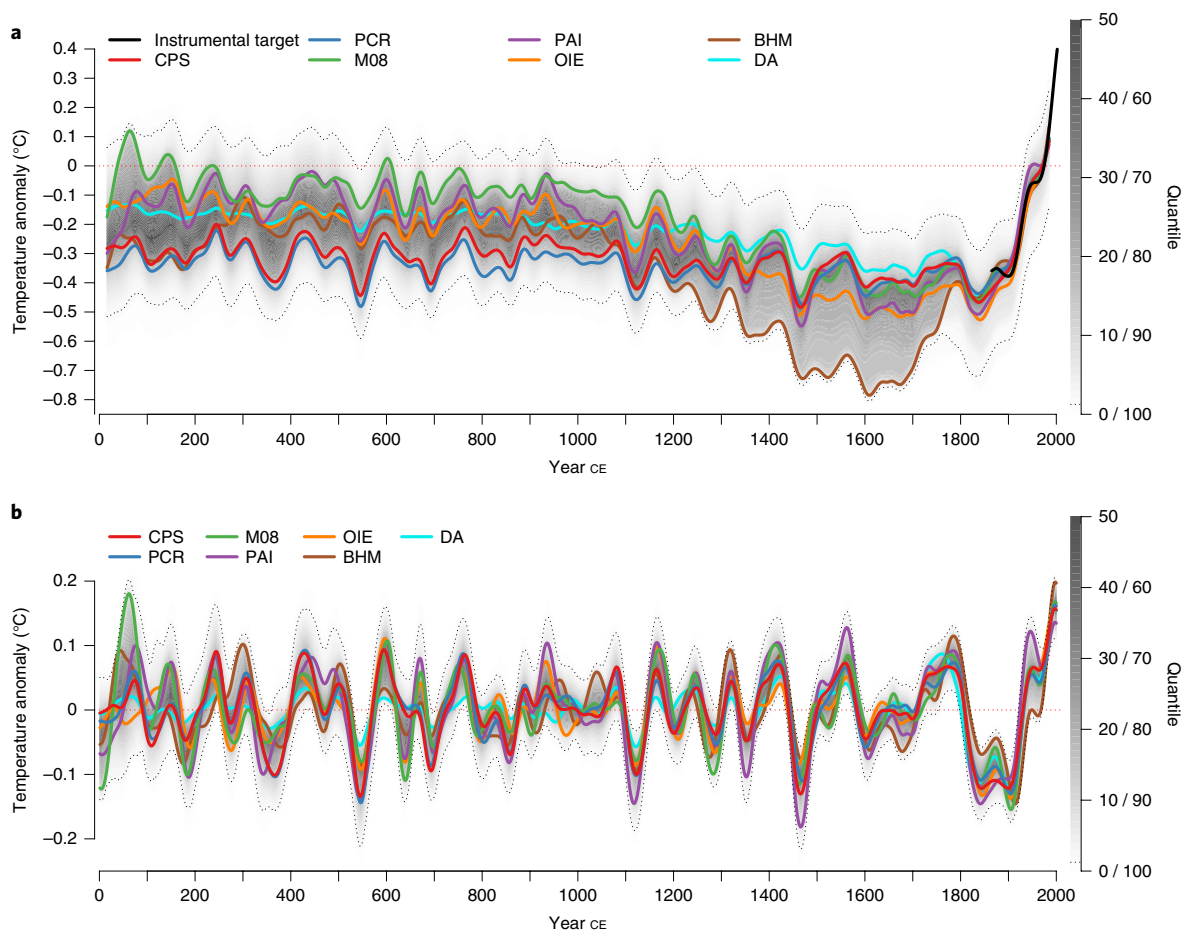


Fig. 1 | Global mean surface temperature history over the Common Era. a, Temperature anomalies with respect to 1961–1990 CE. The coloured lines represent 30 yr low-pass-filtered ensemble medians for the individual reconstruction methods. The grey shading shows the quantiles of all reconstruction ensemble members from all seven methods (right-hand y axis); the 2.5th and 97.5th percentiles are indicated with black dotted lines. The black curve is instrumental data²⁵ for 1850–2017 CE. **b,** Same as **a**, but for the 30 to 200 yr bandpass-filtered ensemble; instrumental data are not shown.

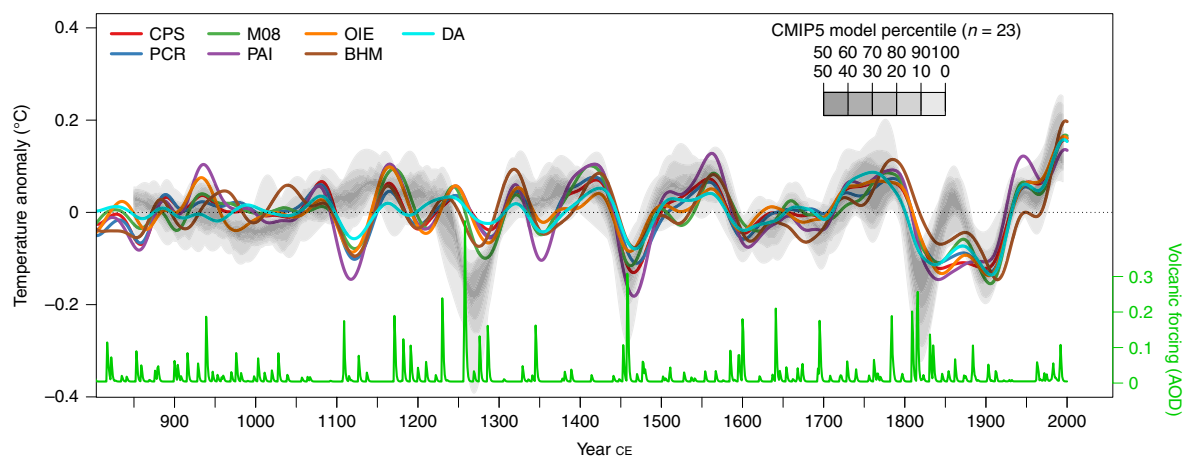


Fig. 2 | MDV in reconstructions and models and volcanic forcing over the past millennium. Temperature anomalies with respect to 1916–1995 CE are shown. Coloured lines are the ensemble median reconstructions from the different methods subjected to 30 to 200 yr bandpass filters. The grey shading shows the model simulation percentiles. The green curve (right-hand y axis) shows the volcanic forcing⁵².

means of three different CESM-LME model experiments (volcanic, solar and greenhouse gas (GHG) only) onto the reconstructions (see equation (3) in the Methods). Choosing combinations of model experiments that account for different forcings allows us

to estimate the relative contribution associated with each forcing (Fig. 3a, right). Results show that solar forcing is not detectable at multidecadal timescales in the reconstructions (scaling factors not significantly above zero), but volcanic and GHG forcing responses

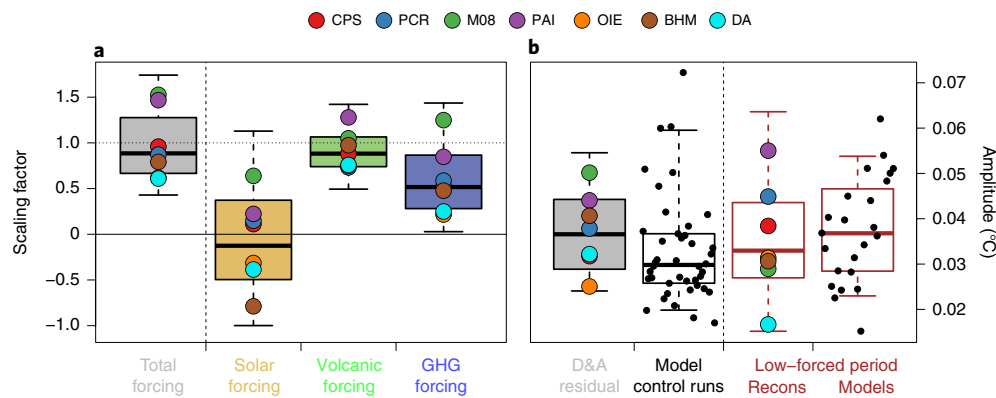


Fig. 3 | Pre-industrial forcing response and magnitude of unforced MDV. a, Multidecadal detection and attribution scaling factors over 1300–1800 CE for the CESM1-CAM5 model based on differently forced runs. The grey box and whisker plot shows the D&A experiment using fully forced runs. The yellow, green and blue box and whisker plots show multivariate experiments combining runs with solar forcing only, volcanic forcing only and GHG forcing only, respectively. Circles represent the median of the individual reconstruction methods. **b**, Estimates of unforced natural variability based on the D&A regression residuals of the full-forcing experiment (grey), the standard deviation of the pre-industrial control simulations (black) and reconstruction and model standard deviation during the period of low external forcing variability (850–1100 CE; brown, for details see Methods). Boxes represent the interquartile range and whiskers represent the 90% range. Black points represent the individual model simulations.

are consistent in pre-industrial multidecadal reconstructions and simulations⁴⁵. Volcanic forcing has scaling factors very close to one, confirming the agreement illustrated in Fig. 2. A weak relationship between solar forcing and GMST is confirmed by cross-wavelet analysis (Supplementary Fig. 16), suggesting that solar variability as currently reconstructed cannot account for GMST variability over the Common Era, although it has been detected on multidecadal timescales regionally^{46,47} as well as over the Northern Hemisphere on multicentennial timescales⁴⁷.

The residuals of the D&A analysis can be used to quantitatively estimate the magnitude of internal unforced MDV in the reconstructions (Methods). The residuals of the total forcing D&A analysis (Fig. 3a, leftmost box and whisker plot) are shown in Fig. 3b, and they are interpretable as the unforced variability in the actual climate system after the linear response to total forcing simulations is removed. This estimate of the unforced variability in the actual climate system is consistent with the variability in pre-industrial control simulations (Fig. 3b, left two box and whisker plots): 99% of estimates based on the D&A residuals are within the 95% range of control simulations. The interquartile range of all possible paired comparisons also encompasses zero (Supplementary Fig. 8).

Differences in the estimates from the two lines of evidence may arise from both random and systematic uncertainties in both reconstructions and simulations⁴⁸ (equation (3) in the Methods). Reconstruction uncertainties include underestimation of the reconstructed variance due to sparse observations, temporal smoothing, integration or seasonal bias, structural data model error and large observational errors^{36,49–51}. Simulation uncertainties arise from errors in amplitude or timing of estimated radiative forcings⁵², incorrect timescales of the simulated responses to forcing⁵³, and underestimation of the forced and unforced response because of an incomplete representation of processes in the actual climate system⁵⁴.

To evaluate the likelihood of these biases influencing our results, we can compare the variability in reconstructions and model simulations for the Medieval ‘quiet’ period⁵⁵ for external radiative forcing (850–1100 CE), which is independent from the period used for D&A (Fig. 3b, rightmost two box and whisker plots). This analysis suggests that there is no significant difference between reconstructed and simulated unforced variability (Methods). Our analysis does not confirm earlier findings of models underestimating

internal MDV^{15,56}. Instead, it suggests that current models are able to realistically simulate the relative magnitudes of the externally and internally forced GMST on multidecadal timescales.

Long-term context of the recent warming rates

To place recent warming rates into a long-term context, we calculate 51 yr running GMST trends from the unfiltered reconstructions (Fig. 4a). The different reconstruction methods agree on the magnitudes of multidecadal trends over the Common Era, as shown by the narrow uncertainty range in Fig. 4a.

Energy balance model (EBM, green line in Fig. 4; see also Supplementary Figs. 14 and 15) simulations with volcanic forcing^{52,57} suggest a multidecadal temperature response to secular frequency and amplitude variations in the forcing (Supplementary Figs. 12 and 13), which is qualitatively consistent with reconstructions. The fractional multidecadal variance increases with increasing frequency of volcanic eruptions (Supplementary Fig. 15). Volcanic eruptions are found to coincide with strong multidecadal cooling trends in both reconstructions and model simulations⁵⁸, and this is then followed by an increased probability of strong warming trends, due to the recovery from the volcanic cooling⁵⁹ (Fig. 4a and Supplementary Figs. 9–13). Both the cooling and subsequent warming trends are found to occur much more frequently than would be expected by chance.

Reconstructions and models yield practically identical values for the upper range of pre-industrial temperature trends (horizontal lines in Fig. 4a), again suggesting that climate models are skilful in simulating the natural range of multidecadal GMST variability.

Of the reconstruction ensemble members, 79% have the largest 51 yr trend within the twentieth century, which includes two distinct periods of large trends. The first reflects the early twentieth century warming⁶⁰, which was shown to originate from a combination of forcings including anthropogenic forcings and internal multidecadal variability of the climate system³. The second reflects the modern period of strong warming, which extends from the mid-1970s to today. The temperature trends during these two industrial-era periods are outside the range of pre-industrial variability, even more so when compared with control runs, in which strong warming trends after volcanic cooling do not occur (dashed red line in Fig. 4a). All instrumental 51 yr trends starting in 1948 CE (and ending 1998 CE) or later exceed the 99th percentile of reconstructed pre-industrial 51 yr trends.

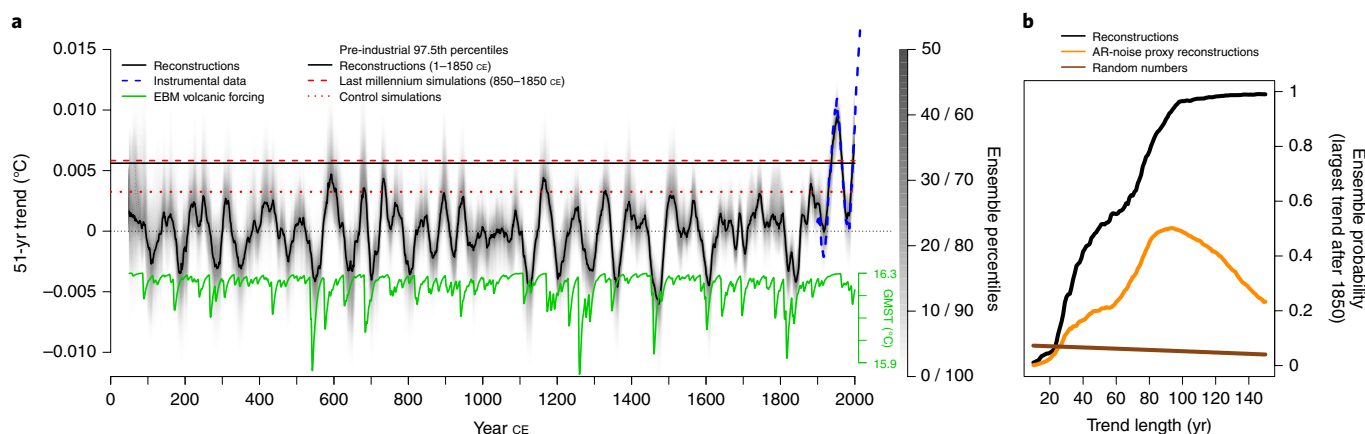


Fig. 4 | Multidecadal temperature trends over the Common Era. a, 51-year trends in reconstructions (ensemble percentiles shaded) and instrumental data²⁵, with the temperature response to volcanic forcing³² based on an EBM. For the trends, the years on the x axis represent the end year of the 51-yr trends. **b**, Ensemble probability of the largest trend occurring after 1850 CE as a function of trend length.

The extraordinary rate of the industrial-era temperature increase is evident on timescales longer than approximately 20 yr (Fig. 4b), for which the probabilities of the occurrence of the largest trends greatly exceed the values expected from chance alone, using random noise predictors with realistic memory properties (Methods).

In addition to using a variety of reconstruction methods, we assessed the robustness of our analysis of multidecadal GMST variability and trends using a range of sensitivity tests, including reconstructions based on detrended calibration and different realizations of noise proxies, adjustments for variance changes back in time, and different calibration periods and proxy subsets. Our inferences on the multidecadal GMST variability for the Common Era are robust to all these permutations (Supplementary Figs. 17–20). Nevertheless, we cannot rule out biases due to errors in the individual proxy records and the unequal spatiotemporal distribution of proxy data (Supplementary Fig. 1). Warm-season-sensitive records from the Northern Hemisphere high and mid latitudes dominate the collection of proxy records²¹, thus our results may be biased towards this region and season, although these effects are reduced by using the global mean (Supplementary Section 3). The coherency across reconstruction methods nevertheless suggests a reduced sensitivity of the results to methodological choices compared with earlier attempts²³, perhaps in large part because of the use of the largest database of temperature-sensitive proxy data yet assembled²¹. With respect to the forced climate simulations, independent uncertainties include: the specification of long-timescale physical processes; relatively poorly known, slowly varying forcings; and representation of grid- and sub-grid-scale processes⁶¹. That the reconstructions presented here are coherently phased and of similar amplitude to CMIP5 simulated MDV suggests improved confidence in both these sources of information.

Agreement on both the timing and amplitude of GMST variability across the reconstruction and simulation ensembles suggests that some aspect of the observed multidecadal variability is externally forced, principally by changes in the frequency and amplitude of volcanic forcing over the pre-industrial past millennium^{31,45,62} and anthropogenic GHGs and aerosols thereafter⁶³ (Supplementary Fig. 7c). Furthermore, we find that estimates of the unforced variability from reconstructions and simulations span similar ranges to independent uncertainties. Our results suggest that multimodel ensemble projections for the near term produce reasonable estimates of the spread in global mean temperature change associated with both forced and unforced MDV, further justifying the use of such models as a basis for regional- to

global-scale simulations⁶⁴ and studies of associated societal impact, adaptation and resilience⁶⁵.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at <https://doi.org/10.1038/s41561-019-0400-0>.

Received: 6 November 2018; Accepted: 5 June 2019;

Published online: 24 July 2019

References

- Masson-Delmotte, V. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) 383–464 (IPCC, Cambridge Univ. Press, 2013).
- Abram, N. J. et al. Early onset of industrial-era warming across the oceans and continents. *Nature* **536**, 411–418 (2016).
- Hegerl, G. C., Brönnimann, S., Schurer, A. & Cowan, T. The early 20th century warming: anomalies, causes, and consequences. *WIREs Clim. Change* **9**, e522 (2018).
- Medhaug, I., Stolpe, M. B., Fischer, E. M. & Knutti, R. Reconciling controversies about the ‘global warming hiatus’. *Nature* **545**, 41–47 (2017).
- Deser, C. & Phillips, A. An overview of decadal-scale sea surface temperature variability in the observational record. *PAGES Mag.* **25**, 2–6 (2017).
- Keenlyside, N. S., Latif, M., Jungclauss, J., Kornbluh, L. & Roeckner, E. Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature* **453**, 84–88 (2008).
- Stott, P. A. et al. External control of 20th century temperature by natural and anthropogenic forcings. *Science* **290**, 2133–2137 (2000).
- Deser, C., Knutti, R., Solomon, S. & Phillips, A. S. Communication of the role of natural variability in future North American climate. *Nat. Clim. Change* **2**, 775–779 (2012).
- Hawkins, E. & Sutton, R. The potential to narrow uncertainty in regional climate predictions. *Bull. Am. Meteorol. Soc.* **90**, 1095–1108 (2009).
- Cassou, C. et al. Decadal climate variability and predictability: challenges and opportunities. *Bull. Am. Meteorol. Soc.* **99**, 479–490 (2018).
- Santer, B. D. et al. Causes of differences in model and satellite tropospheric warming rates. *Nat. Geosci.* **10**, 478–485 (2017).
- Henley, B. J. et al. Spatial and temporal agreement in climate model simulations of the Interdecadal Pacific Oscillation. *Environ. Res. Lett.* **12**, 044011 (2017).
- Kajtar, J. B. et al. Global mean surface temperature response to large-scale patterns of variability in observations and CMIP5. *Geophys. Res. Lett.* **46**, 2232–2241 (2019).
- Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
- Laepfle, T. & Huybers, P. Ocean surface temperature variability: large model–data differences at decadal and longer periods. *Proc. Natl Acad. Sci. USA* **111**, 16682–16687 (2014).

16. Rehfeld, K., Münch, T., Ho, S. L. & Laepple, T. Global patterns of declining temperature variability from the last glacial maximum to the holocene. *Nature* **554**, 356–359 (2018).
17. Zhu, F. et al. Climate models can correctly simulate the continuum of global-average temperature variability. *Proc. Natl Acad. Sci. USA* **116**, 8728–8733 (2019).
18. Ding, Y. et al. Ocean response to volcanic eruptions in Coupled Model Intercomparison Project 5 simulations. *J. Geophys. Res. Oceans* **119**, 5622–5637 (2014).
19. 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. *J. Geophys. Res.* **117**, D08101 (2012).
20. Meehl, G. A. et al. Decadal prediction. *Bull. Am. Meteorol. Soc.* **90**, 1467–1486 (2009).
21. PAGES2k Consortium. A global multiproxy database for temperature reconstructions of the Common Era. *Sci. Data* **4**, 170088 (2017).
22. Christiansen, B. & Ljungqvist, F. C. Challenges and perspectives for large-scale temperature reconstructions of the past two millennia. *Rev. Geophys.* **55**, 40–96 (2017).
23. 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. *Geophys. Res. Lett.* **42**, 7162–7170 (2015).
24. Smerdon, J. E. & Pollack, H. N. Reconstructing Earth's surface temperature over the past 2000 years: the science behind the headlines. *WIREs Clim. Change* **7**, 746–771 (2016).
25. Cowtan, K. & Way, R. G. Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Q. J. R. Meteorol. Soc.* **140**, 1935–1944 (2014).
26. Rehfeld, K., Trachsel, M., Telford, R. J. & Laepple, T. Assessing performance and seasonal bias of pollen-based climate reconstructions in a perfect model world. *Clim. Past* **12**, 2255–2270 (2016).
27. Ljungqvist, F. C., Krusic, P. J., Brattström, G. & Sundqvist, H. S. Northern Hemisphere temperature patterns in the last 12 centuries. *Clim. Past* **8**, 227–249 (2012).
28. Esper, J. et al. Ranking of tree-ring based temperature reconstructions of the past millennium. *Quat. Sci. Rev.* **145**, 134–151 (2016).
29. Esper, J., Cook, E. R., Krusic, P. J., Peters, K. & Schweingruber, F. H. Tests of the RCS method for preserving low-frequency variability in long tree-ring chronologies. *Tree-Ring Res.* **59**, 81–98 (2003).
30. Klippel, L., George, S. S., Büntgen, U., Krusic, P. J. & Esper, J. Differing pre-industrial cooling trends between tree-rings and lower-resolution temperature proxies. *Clim. Past Discuss.* <https://doi.org/10.5194/cp-2019-41> (2019).
31. McGregor, H. V. et al. Robust global ocean cooling trend for the pre-industrial Common Era. *Nat. Geosci.* **8**, 671–677 (2015).
32. St. George, S. An overview of tree-ring width records across the Northern Hemisphere. *Quat. Sci. Rev.* **95**, 132–150 (2014).
33. Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M. & Anchukaitis, K. N. Applications of proxy system modeling in high resolution paleoclimatology. *Quat. Sci. Rev.* **76**, 16–28 (2013).
34. Babst, F. et al. Twentieth century redistribution in climatic drivers of global tree growth. *Sci. Adv.* **5**, eaat4313 (2019).
35. Smerdon, J. E. Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments. *WIREs Clim. Change* **3**, 63–77 (2012).
36. Wang, J., Emile-Geay, J., Guillot, D., Smerdon, J. E. & Rajaratnam, B. Evaluating climate field reconstruction techniques using improved emulations of real-world conditions. *Clim. Past* **10**, 1–19 (2014).
37. Toohey, M. & Sigl, M. Volcanic stratospheric sulfur injections and aerosol optical depth from 500 BCE to 1900 CE. *Earth Syst. Sci. Data* **9**, 809–831 (2017).
38. Crowley, T. J. & Unterman, M. B. Technical details concerning development of a 1200 yr proxy index for global volcanism. *Earth Syst. Sci. Data* **5**, 187–197 (2013).
39. Gao, C., Robock, A. & Ammann, C. Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models. *J. Geophys. Res.* **113**, D23111 (2008).
40. Marotzke, J. & Forster, P. M. Forcing, feedback and internal variability in global temperature trends. *Nature* **517**, 565–570 (2015).
41. Allen, M. R. & Stott, P. A. Estimating signal amplitudes in optimal fingerprinting, part I: theory. *Clim. Dynam.* **21**, 477–491 (2003).
42. Bindoff, N. L. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) 867–952 (IPCC, Cambridge Univ. Press, 2013).
43. 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. *J. Clim.* **26**, 6954–6973 (2013).
44. Otto-Bliesner, B. L. et al. Climate variability and change since 850 CE: an ensemble approach with the community earth system model. *Bull. Am. Meteorol. Soc.* **97**, 735–754 (2016).
45. Schurer, A. P., Tett, S. F. B. & Hegerl, G. C. Small influence of solar variability on climate over the past millennium. *Nat. Geosci.* **7**, 104–108 (2013).
46. Taricco, C., Mancuso, S., Ljungqvist, F. C., Alessio, S. & Ghil, M. Multispectral analysis of Northern Hemisphere temperature records over the last five millennia. *Clim. Dynam.* **45**, 83–104 (2015).
47. Anchukaitis, K. J. et al. Last millennium Northern Hemisphere summer temperatures from tree rings: part II, spatially resolved reconstructions. *Quat. Sci. Rev.* **163**, 1–22 (2017).
48. PAGES2k-PMIP3 group. Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium. *Clim. Past* **11**, 1673–1699 (2015).
49. Frost, C. & Thompson, S. G. Correcting for regression dilution bias: comparison of methods for a single predictor variable. *J. R. Stat. Soc. Ser. A* **163**, 173–189 (2000).
50. von Storch, H. Reconstructing past climate from noisy data. *Science* **306**, 679–682 (2004).
51. 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. *Sci. Rep.* **8**, 7572 (2018).
52. Sigl, M. et al. Timing and climate forcing of volcanic eruptions for the past 2,500 years. *Nature* **523**, 543–549 (2015).
53. Jungclaus, J. H. et al. The PMIP4 contribution to CMIP6—part 3: the last millennium, scientific objective, and experimental design for the PMIP4 past1000 simulations. *Geosci. Model Dev.* **10**, 4005–4033 (2017).
54. Diffenbaugh, N. S., Pal, J. S., Trapp, R. J. & Giorgi, F. Fine-scale processes regulate the response of extreme events to global climate change. *Proc. Natl Acad. Sci. USA* **104**, 15774–15778 (2005).
55. Bradley, R. S., Wanner, H. & Diaz, H. F. The medieval quiet period. *Holocene* **26**, 990–993 (2016).
56. Neukom, R. et al. Inter-hemispheric temperature variability over the past millennium. *Nat. Clim. Change* **4**, 362–367 (2014).
57. Goosse, H. *Climate System Dynamics and Modelling* (Cambridge Univ. Press, 2015).
58. Miller, G. H. et al. Abrupt onset of the little ice age triggered by volcanism and sustained by sea-ice/ocean feedbacks. *Geophys. Res. Lett.* **39**, L02708 (2012).
59. Brönnimann, S. et al. Last phase of the Little Ice Age forced by volcanic eruptions. *Nat. Geosci.* <https://doi.org/10.1038/s41561-019-0402-y> (2019).
60. Brönnimann, S. Early twentieth-century warming. *Nat. Geosci.* **2**, 735–736 (2009).
61. Flato, G. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) 741–866 (IPCC, Cambridge Univ. Press, 2013).
62. Atwood, A. R., Wu, E., Frierson, D. M. W., Battisti, D. S. & Sachs, J. P. Quantifying climate forcings and feedbacks over the last millennium in the CMIP5-PMIP3 models. *J. Clim.* **29**, 1161–1178 (2016).
63. IPCC *Climate Change 2013: The Physical Science Basis* (Cambridge Univ. Press, 2013).
64. Giorgi, F. & Gao, X.-J. Regional earth system modeling: review and future directions. *Atmos. Ocean Sci. Lett.* **11**, 189–197 (2018).
65. Seneviratne, S. I. et al. The many possible climates from the Paris Agreement's aim of 1.5°C warming. *Nature* **558**, 41–49 (2018).
66. Mann, M. E. et al. Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc. Natl Acad. Sci. USA* **105**, 13252–13257 (2008).
67. Luterbacher, J. et al. Reconstruction of sea level pressure fields over the Eastern North Atlantic and Europe back to 1500. *Clim. Dynam.* **18**, 545–561 (2002).
68. Shi, F., Zhao, S., Guo, Z., Goosse, H. & Yin, Q. Multi-proxy reconstructions of May–September precipitation field in China over the past 500 years. *Clim. Past* **13**, 1919–1938 (2017).
69. Hanhijärvi, S., Tingley, M. P. & Korhola, A. Pairwise comparisons to reconstruct mean temperature in the Arctic Atlantic Region over the last 2,000 years. *Clim. Dynam.* **41**, 2039–2060 (2013).
70. Barboza, L., Li, B., Tingley, M. P. & Viens, F. G. Reconstructing past temperatures from natural proxies and estimated climate forcings using short- and long-memory models. *Ann. Appl. Stat.* **8**, 1966–2001 (2014).
71. Hakim, G. J. et al. The last millennium climate reanalysis project: framework and first results. *J. Geophys. Res. Atmos.* **121**, 6745–6764 (2016).

Acknowledgements

This is a contribution to the PAGES 2k Network. PAGES is supported by the US National Science Foundation and the Swiss Academy of Sciences. PAGES 2k Network members are acknowledged for providing input proxy data. Some calculations were run on the Ubelix cluster at the University of Bern. S. Hanhijärvi provided the PAI code. M. Grosjean, S. J. Phipps and J. Werner provided inputs at different stages of the project. R.N. is supported by Swiss NSF grant number PZ00P2_154802. K.R. is funded by DFG grant number RE3994-2/1. S.B. acknowledges funding from the European Union (project 787574). E.S. is funded by the NSFC (grants numbers 41877440; 41430531; 41690114). A.S. was supported by NERC under the Belmont forum, Grant PacMedy (grant number

NE/P006752/1). B.J.H. acknowledges funding from the Australian Research Council, Melbourne Water and DELWP on Linkage Project (LP150100062) and support from the Australian Bureau of Meteorology. B.J.H. also acknowledges support from the ARC Centre of Excellence for Climate Extremes (CE170100023).

Author contributions

R.N. coordinated the project. R.N. and J.E.-G. provided and generated input data. R.N. (PCR, CPS, PAI), F.S. (OIE, M08), M.P.E. (DA) and L.A.B. (BHM) developed and performed the indicated GMST reconstructions. R.N., K.R. and M.N.E. analysed reconstruction results. L.L. and A.S. performed the D&A analysis. F.Z. calculated the solar cross-wavelet analysis. K.R. performed the EBM analyses. J.F. and V.V. contributed to other data analysis. R.N. made the figures. R.N., D.S.K., M.N.E. and K.R. wrote the paper. L.A.B., S.B., J.E.-G., M.P.E., M.N.E., J.F., G.J.H., B.J.H., D.S.K., F.C.L., R.N., N.M., K.R., A.S., F.S. and L.v.G. designed the study, discussed the results and contributed to the writing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41561-019-0400-0>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to R.N.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

PAGES 2k Consortium

Raphael Neukom^{1*}, Luis A. Barboza², Michael P. Erb³, Feng Shi^{4,5}, Julien Emile-Geay⁶, Michael N. Evans⁷, Jörg Franke¹, Darrell S. Kaufman³, Lucie Lücke⁸, Kira Rehfeld^{9,10}, Andrew Schurer⁸, Feng Zhu⁶, Stefan Brönnimann¹, Gregory J. Hakim¹¹, Benjamin J. Henley^{12,13,14}, Fredrik Charpentier Ljungqvist^{15,16,17}, Nicholas McKay³, Veronika Valler¹ and Lucien von Gunten¹⁸

¹Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, Bern, Switzerland. ²Escuela de Matematica-CIMPA, Universidad de Costa Rica, San Jose, Costa Rica. ³School of Earth and Sustainability, Northern Arizona University, Flagstaff, AZ, USA. ⁴Key Laboratory of Cenozoic Geology and Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China. ⁵CAS Center for Excellence in Life and Paleoenvironment, Beijing, China. ⁶Department of Earth Sciences, University of Southern California, Los Angeles, CA, USA. ⁷Department of Geology and ESSIC, University of Maryland, College Park, MD, USA. ⁸School of Geosciences, University of Edinburgh, Edinburgh, UK. ⁹British Antarctic Survey, Cambridge, UK. ¹⁰Institute of Environmental Physics, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany. ¹¹Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA. ¹²School of Earth Sciences, University of Melbourne, Melbourne, Victoria, Australia. ¹³School of Earth, Atmosphere and Environment, Monash University, Clayton, Victoria, Australia. ¹⁴ARC Centre of Excellence for Climate Extremes, University of Melbourne, Melbourne, Victoria, Australia. ¹⁵Department of History, Stockholm University, Stockholm, Sweden. ¹⁶Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden. ¹⁷Department of Geography, University of Cambridge, Cambridge, UK. ¹⁸PAGES International Project Office, Bern, Switzerland. *e-mail: neukom@giub.unibe.ch

Methods

Input data and preprocessing. *Palaeoclimate observations.* We use version 2.0.0 of the global PAGES 2k temperature proxy data collection²¹. For the reconstructions presented in the main text, we use the subset of records selected on the basis of regional temperature screening and to account for false discovery rates (R-FDR subset²¹). This screening reduces the total number of records from 692 to 257, but increases the GMST reconstruction skill for most methods and skill metrics. A comparison of reconstructions and evaluation statistics of the full versus screened networks is provided in Supplementary Fig. 17. Of the screened proxy records, 210 are annually resolved and are included by all reconstruction methods. Sub-annually resolved time series (29 records) were averaged over the April–March annual window. The 47 records with lower-than-annual resolution are only included by the M08, OIE and PAI methods (see below). Although none of the methods explicitly consider age uncertainties, the impact of age uncertainty on reconstructions can be estimated by sampling from a set of age ensembles for each of the age-uncertain records (for example, ref. ⁷²). However, only a subset (15%) of the age-uncertain records in version 2.0.0 of the PAGES 2k database include age ensembles or the requisite geochronological data to generate new age ensembles. Consequently, full consideration of age uncertainty is outside the scope of this research; nevertheless, this probably has minimal impacts on the reconstructions, as the final proxy network mainly consists of records with very small to negligible age uncertainties (81% of records are from tree or coral archives).

Instrumental observations. We use the Cowtan and Way²⁵ global mean temperature dataset, which is based on a kriging-interpolated version of HadCRUT4¹⁹. We use an annual average based on the April–March window that represents the ‘tropical year’, which has the advantage of not interrupting the growing season of trees in both the Northern and Southern Hemispheres.

Experiment overview and settings. The time window 1850–2000 CE was used for calibration. For evaluation, we ran a second experiment calibrating over 1916–1995 CE and validating over 1881–1915 CE. Evaluation graphs are provided in the Supplementary Section 2. Results based on this shorter calibration period, which encompasses a period of significantly improved spatiotemporal coverage with instrumental and proxy data, are presented in Supplementary Fig. 17. All reconstruction methods generated an ensemble of 1,000 GMST reconstructions, which was used for uncertainty quantification.

Reconstruction methods. Details on each method are provided in the following subsections. Table 1 summarizes some key aspects and references for the methods. Methodological constraints required some minor adaptations to the experimental set-up for some techniques. These are as follows: CPS, PCR, BHM and DA include only annually and more finely resolved proxy records. CPS and PCR exclude records with more than 33% missing values and infill the remaining missing values over the calibration period. DA generates GMST from a spatially resolved temperature reconstruction. Multiple approaches to weighting the different reconstruction methods to generate a best estimate GMST time series are presented and discussed in Supplementary Section 4.

CPS. One of the most basic methods for quantitative climate index reconstruction, CPS has been applied widely for local- to global-scale reconstructions of temperature and other variables^{66,73–76}. CPS generates reconstructions using two steps. First, the proxy records are converted into a single time series (composite) by calculating a weighted mean. Second, this composite is scaled to the mean and standard deviation of the reconstruction target over the calibration period. Here, we use a similar implementation of CPS as ref. ⁵⁶: the proxies are weighted by their non-detrended correlation with the GMST reconstruction target over the calibration period. A nested approach is applied, such that each year’s CPS reconstruction is based on a calibration using all proxies that are available in that year. Several key parameters are resampled to quantify uncertainties for each member of the 1,000-member ensemble:

- 10% of proxy records are withheld from calibration.
- The weight of each proxy record (target correlation r) is multiplied by a factor between 1/1.5 and 1.5. In contrast to other studies (such as ref. ⁷⁷), we do not perturb r with an exponent but use a factor centred at 1 to ensure that the reconstruction ensemble is centred around a reconstruction using non-perturbed weighting.

In addition, each ensemble member is perturbed with red noise that has the same standard deviation and first-order autoregression (AR(1)) coefficient as its residuals from the reconstruction target over the calibration period. We use an AR(1) model because the short length of time available for model fitting and evaluation does not allow for a reliable estimate of coefficients for higher-order and more complex models. Furthermore, CPS residuals only exhibit significant autocorrelation at lag 1 (not shown). Thus, the total uncertainty expressed by the ensemble spread is based on both parameter and calibration uncertainty⁷⁸. For a detailed discussion of this ensemble approach refer to refs. ^{56,78}.

As for many linear methods, CPS is prone to variance losses in the reconstructions and a bias towards zero. Both the bias and reconstruction variance

are directly related to the correlation of the proxies with the reconstruction target²². CPS is computationally efficient and the quality (in terms of skill measures, for example) of its results is often comparable to more advanced (and often more expensive and parameter-dependent) methods (see Supplementary Figs. 2, 3 and refs. ^{56,78,79}).

For proxy records with lower-than-annual nominal resolution, the correlations with the instrumental target have considerably reduced degrees of freedom. Furthermore many of these records are not interpolated to a value for each year in the PAGES 2k database v2.0.0, thus cannot be directly used for CPS. We therefore use only records with a median resolution (‘ResMed’ variable in the PAGES 2k proxy database) not exceeding 1. Furthermore, we only use records with a maximum of 1/3 of missing data within the calibration interval. This yields a total of 210 proxy records within the R-FDR screened subset. Missing data in the calibration window (2.23%) are infilled using DINEOF⁸⁰.

PCR. This is a linear method and, along with CPS and RegEM (see below), is one of the most commonly used methods to reconstruct climate indices and fields over the Common Era^{22,24,75,81,82}. PCR reduces the dimensionality of the proxy matrix by using only a certain number of its principal components as predictors in the regression. PCR has similar advantages and deficiencies to CPS. There is a large literature describing potential issues of the method and its parameter choices^{22,24,50,83–87}.

Here, we use the nested PCR approach originally presented by ref. ⁶⁷ and implemented in many reconstructions of regional to hemispheric climate^{56,79,88–92}. Our implementation is similar to ref. ⁵⁶; we use ordinary least squares to generate nested 1,000-member ensemble reconstructions. We use identical parameters and input data to those described for CPS above with a few exceptions: as additional parameter perturbation, principal components are truncated so that they explain between 40% and 90% of the total variance in the proxy matrix. Similar to the weighting in CPS, each proxy record is multiplied by a factor between 1/1.5 and 1.5 to perturb its weight within the PCA routine. All records are scaled to a mean of zero and unit variance over their common period of overlap before this variance adjustment.

M08. The core of the regularized errors-in-variables approach⁶⁶ is the regularized expectation maximization (RegEM⁹³) algorithm with truncated total least squares (TLS) regression⁹⁴. The reconstruction by ref. ⁶⁶ (hereafter M08) is the most widely used GMST reconstruction over the Common Era in the recent literature^{36,66,79,86,95,96} and has been much criticized²². Here we use the M08 implementation of RegEM to allow a direct comparison. Compared with other methods used herein, the most distinctive feature is a hybrid process to decompose two levels with high/low-frequency bands of the reconstructions using the RegEM method with the truncated TLS regression. Thus, the method is able to combine low- and high-frequency proxies. The instrumental and gridded April–March average temperatures²⁵ were interpolated using RegEM to cover the whole period (1850–2000 CE) and used to calculate the regional properties of each record; for example, the correlation coefficient between the proxy record and the regional and instrumental data during the calibration period.

The uncertainty in the original M08 method was calculated using the standard deviation of instrumental data during the calibration period and the correlation coefficient between the proxy reconstruction and instrumental data during the evaluation period⁶⁶. Here, we adapted the uncertainty estimation to the ensemble reconstruction by perturbing each ensemble member with red noise of the same standard deviation and AR(1) coefficient as the residuals of the reconstruction during the evaluation period^{56,78}.

PAI. Pairwise comparison is a nonlinear method⁶⁹ that allows proxy records with different temporal resolution to be combined. It has been used in a number of continental-scale temperature reconstructions^{69,78,97}. PAI generates a nonlinear, but monotonically increasing function to reconstruct climate by performing pairwise comparisons of the proxy records. The order of successive data points in each proxy time series is compared and the relative agreement of this ordering for all proxies is then calculated. The strength of this agreement is then used to predict temporal changes in the target climatic variable. Similar to CPS, PAI generates a unitless composite on the basis of the pairwise comparisons. In contrast to CPS, which makes use of target correlations to generate the composite, the composite derived by PAI is fully independent of the reconstruction target. Therefore, careful proxy selection is important for PAI to yield skilful results. The composite is scaled to the reconstruction target over the calibration interval. Reconstruction ensembles are generated by calculating bootstraps, resampling the proxy records and adding noise to the target data. A detailed description of the method is provided in ref. ⁶⁹. Here we use the same parameters as in ref. ⁹⁷.

The key advantage of PAI compared to most other methods used at present is its ability to combine high- and low-resolution proxy datasets, including records with irregular temporal resolution or those that do not cover the calibration period.

OIE. The OIE method was derived from the CPS method, and developed with several versions inspired by the local (LOC) method⁹⁸, a Bayesian framework (BARCAST)⁹⁹, the generalized likelihood uncertainty estimation method^{100,101} and the ensemble reconstructions⁵⁶.

We use version 2.0 of the OIE method, which was developed to reconstruct the GMST by combining previously published reconstructions (four Northern Hemisphere reconstructions with annual resolution and covering the past two millennia^{66,102–104} and a Southern Hemisphere ensemble reconstruction⁵⁶ using direct and indirect regression, which generated a reconstruction through four steps. First, we matched the variance of each proxy record using the previous hemispheric reconstructions. Second, we averaged the matched records to generate a composited proxy record, which is like the CPS method to weight and average the screened records to a composited record. Third, we regressed the composite using the E-LOC method⁹⁵, which set the regression coefficients as random variables within the ranges of the direct and indirect regressions, and the regression coefficients were determined using the generalized likelihood uncertainty estimation method^{100,101}. Finally, the uncertainty is expressed as the ensemble spread, in which each ensemble member is perturbed with red noise that has the same standard deviation and AR(1) coefficient calculated from the residuals of the reconstruction in the third step^{56,78}.

BHM. The BHM, defined in ref.⁷⁰, is based on two ingredients: a model of the GMST (T) that depends on external forcing and additive noise (equation (1)); and a model of the reduced proxy signal (P) as linearly dependent on T , with additive noise (equation (2)). Specifically:

$$T_t = \beta_0 + \beta_1 S_t + \beta_2 V_t + \beta_3 C_t + \sigma_T \epsilon_t \quad (1)$$

$$P_t = \alpha_0 + \alpha_1 T_t + \sigma_P \eta_t \quad (2)$$

where t is time (1–2000 CE) and

RP_t is the RP series that captures the common signal of the entire proxy dataset. It is constructed through a Lasso Regression¹⁰⁵ between proxies with similar temporal availability and the observed anomalies²³ during the calibration period. This procedure is repeated over 8 250 yr nests with different temporal availability of the proxies, and the fitted values obtained for each case are combined in a single series. The number of records that were used in each nest vary between 21 (earliest nest) and 199 (latest nest).

T_t is the temperature anomalies series.

S_t is the time series of solar irradiance computed from SATIRE-H¹⁰⁶,

V_t is the transformed volcanic forcing from Easy Volcanic Aerosol (EVA) dataset^{37,107}

C_t is the transformed GHG concentrations taken from CMIP6¹⁰⁸,

η_t and ϵ_t are the zero-mean, unit-variance stationary stochastic processes and σ_P and σ_T are constant variance parameters.

This method proceeds by regressing a linear combination of the PAGES 2k data onto global temperature, and modelling the latter as depending on external forcings. All model parameters are assumed random, and their prior distributions are updated through a combination of Gibbs sampling and the Metropolis–Hastings algorithm. This allows us to estimate the latent variable T_t together with all the parameters of the model and thus to obtain samples of its joint posterior distribution given the set of proxy records. This procedure also allows us to obtain prediction intervals for all the parameters and the latent variable of the model. We also considered a particular case of the model for which all the external forcings C_t , V_t and S_t are zero. In this way, the reconstruction of the latent variable was performed using information from the proxies only. For this particular case we defined η_t as white noise and ϵ_t an AR(1) process. When the external forcings are not null, both η_t and ϵ_t are Gaussian white noise processes.

DA. This method fuses proxy observations and climate model output to reconstruct multiple climate fields, while leveraging climate model physics. The DA procedures used in this study were developed for the Last Millennium Reanalysis (LMR) project^{71,109}. The DA method of the LMR reconstructs full fields rather than just global mean temperature, and it holds advantages over competing methods such as linear regression, both frequentist and Bayesian⁷¹. Because it uses precomputed output from a climate model, this approach is known as offline data assimilation. Offline assimilation enables vast improvements in computational cost with modest loss of reconstruction skill^{110,111}. The DA code used in this paper is adapted from an intermediate version of the code being developed in the LMR project.

The DA results in this study use information from two sources: the PAGES 2k v2.0.0 proxy network²¹ and the NCAR Community Climate System Model version 4 (CCSM4) last millennium simulation¹¹². The proxy records provide the temporal variations related to climate at specific locations, while the model results quantify the covariances between different fields and locations in the climate system. In other words, proxies provide temporal variations while the model covariances spread this information through space and to other climate fields in a dynamically consistent way.

The DA methodology works by first drawing 200 random years from the output of the CCSM4 last millennium simulation. This collection of climate states represents an initial guess of the climate before any assimilation takes place, and is referred to as the model prior. This prior is initially the same for every year of

the reconstruction. The DA method gets no information about temporal changes from the model, and also has no knowledge of when given forcings are present in the climate system. It is therefore worthwhile to compare forced responses in the model and the DA reconstruction as we do in Figs. 2–4.

To relate proxy observations to the model states, proxy system models³³ (PSMs) must be used. While the LMR framework allows for nonlinear PSMs, recent work shows that linear PSMs help safeguard against climate model biases¹¹³, which can be sizable. For tree-ring widths, which can be sensitive to temperature, moisture or both, we use a bivariate linear regression to relate ring width to temperature (HadCRUT4¹⁹, spatially interpolated via kriging²⁵) and moisture (precipitation from the Global Precipitation Climatology Centre¹¹⁴). For all other proxy types, we use a univariate linear regression to the temperature dataset alone.

Using these relationships, an initial estimate of each proxy value is computed from the prior. For every year, the difference between each proxy estimate and the actual proxy observation is computed, and a Kalman filter is used to update the climate state so that these differences are reduced. Because the climate system is updated using the model covariances quantified in the prior, the end result is a dynamically consistent estimate of climate-system anomalies informed by the proxy data. To be used in this assimilation, PAGES 2k v2.0.0 records need to exhibit annual (or better) resolution, as well as at least 25 yr of overlap with the instrumental datasets to calibrate the PSMs. For the primary experiment (which uses the R-FDR screened proxies), 229 records were assimilated, from 134 tree-ring chronologies, 68 corals or sclerosponges, 21 glacier ice cores, 5 lake sediments and 1 bivalve.

Model simulations used. We use the past millennium runs of the following CMIP5 model simulations, yielding an ensemble of 23 simulations: BCC-CSM¹¹⁵, CCSM4¹¹², MPI_ESM_P (members 1–3)^{116–118}, CSIROmk3L-1-2¹¹⁹, GISS-E2-R (members 121, 124 and 127)¹²⁰, HadCM3¹³, CESM_1 (members 1–13)¹⁴. Internal variability was generated from 29 CMIP5 pre-industrial control simulations of length 499–1,198 yr, giving 21,427 yr of internal variability (same models as before plus ACCESS1, BNU-ESM, CanESM2, CMCC-CM, CNRM-CM5, FGOALS (g2, s2), FIO-ESM, GFDL (2G, 2M), HadGEM2, INMCM4, IPSL-CM5A-LR, MRI-CGCM3 and NorESM1). All simulations were aggregated annually by April–March GMST and filtered in the same way as the reconstructions.

Post-processing for figures. To generate anomalies for a given reference period, we subtract the temporal average of each method's ensemble mean over this period from all ensemble members of this method. For low- and bandpass filtering, we use butterworth filters from the R-package signal.

Pre-industrial trends. The significance of the linear pre-industrial trends of the unfiltered ensemble median reconstructions for all methods was tested using the Mann–Kendall test ($n=1,850$, two-tailed) and block-bootstrapping (tsboot() function in R, 100 yr blocks) over the period 0–1850. Both tests yield significant ($\alpha=0.05$) negative trends for all reconstruction methods.

Multidecadal GMST and data–model agreement. Multidecadal GMST variability was analysed by applying a 30–200 yr bandpass butterworth filter to all data (Table 1 and Fig. 2). See Supplementary Fig. 18 for an alternative of Fig. 2 using a 20–100 yr bandwidth. To account for the very high autocorrelation strongly reducing degrees of freedom, and non-normal distributions in the 30–200 yr bandpass-filtered reconstruction and model data, we tested the significance of data–model coherence by comparison with random noise time series. We generated a red noise time series for each reconstruction and model ensemble member with the same AR(1) coefficient as the corresponding reconstruction or model data. These noise time series were then subjected to the 30–200 yr bandpass filter and correlated to each other, as was done for the reconstructions and models. The percentages of data versus model correlations that exceed the noise-based correlations are provided in Table 1. The data–model comparisons are calculated over the 1300–2000 CE period, for which uncertainties in all datasets are relatively low (see main text).

Detection and attribution. To determine the extent of forced variability in the reconstructions, a TLS regression of the fingerprint of external forcings on the reconstructions was performed (Fig. 3). A given temperature reconstruction $Y(t)$ decomposes into a linear combination of the m fingerprints $X_i(t)$ for the climatic response to external forcings, and the internal variability $\nu_0(t)$:

$$Y(t) = \sum_{i=1}^m (X_i(t) - \nu_i(t)) \times \beta_i + \nu_0(t) \quad (3)$$

Each fingerprint $X_i(t)$, given by a model simulation, has its own implementation of internal variability $\nu_i(t)$ that is decreased by averaging over an ensemble of multiple simulations.

To determine the scaling factors β_i , we performed a TLS regression, accounting for the presence of noise in both the regressor and regressand. Such a regression is commonly solved using a regularized optimal detection approach, which was introduced by ref.¹²¹, building on earlier work by ref.⁴¹. Optimization focuses

the regression on features of the forced signal that are most distinct from internal variability. Here, we choose not to optimize the signal and instead carry out a simpler non-optimized regression, following ref. ⁴³. As described in ref. ⁴¹, the scaling factors and the noise-reduced fingerprints $\tilde{X}(t)$ and reconstructions $\tilde{Y}(t)$ can be calculated, using a singular-value decomposition, and defined as:

$$\tilde{Y}(t) = Y(t) - \nu_0(t) \quad (4)$$

$$\tilde{X}(t) = \sum_{i=1}^m (X_i(t) - \nu_i(t)) \quad (5)$$

$$\tilde{Z}(t) = [\tilde{Y}(t), \tilde{X}(t)] \quad (6)$$

To account for the uncertainty due to internal variability, we calculated a range of 1,000 scaling factors per reconstruction by superimposing random slices of internal variability sampled from pre-industrial control simulations on the columns of $\tilde{Z}(t)$. Thus we obtained a distribution of 1×10^6 scaling factors per reconstruction method and fingerprint, accounting for uncertainty from both the reconstruction method and internal variability. The effect of a forcing is said to be detectable if a scaling factor to that forcing is found to be significantly greater than zero. A difference in response to forcing in the model simulations and in the reconstructions is found if the estimated 90% scaling factor confidence interval does not encompass one.

Two different analyses were carried out. The first regresses model simulations with all forcings onto the reconstructions over the time period 1300–1800 CE. This tests whether the response to external forcings is detectable (Fig. 3a, left). For this, we used the ensemble mean of the 13 CESM-LME simulations. Our results are robust to using the CESM model only or the full multimodel mean of all 23 simulations (Supplementary Fig. 7). The second analysis uses a linear combination of the fingerprints of volcanic forcing, solar forcing and GHG forcings, which were generated from the ensemble mean of CESM-LME simulations, regressed onto the reconstructions. This analysis disentangles the forced response into these three different causes (Fig. 3a, right).

The smallest eigenvalue of the singular-value decomposition can be used as an estimate of the sum-of-squares residual of the regression, which is compared, in Fig. 3b, to the model internal variability calculated as the standard deviation of the 500 yr slices of pre-industrial control simulations. As a further estimate of unforced multidecadal GMST variability, we use the 30–200 yr bandpass filtered reconstructions and model simulations during the period 850–1100 CE (brown boxplots in Fig. 3b), which reflects a quiet period of relatively little forcing activity (that is, few large volcanic eruptions and relatively constant solar forcing, see also ref. ⁵⁵).

To test the consistency of estimates of the unforced variance (Fig. 3b, left, right), we applied the following procedures. We first quantified the fraction of estimates based on the D&A residuals that lie within the confidence range of the estimate from control simulations. Note that the estimates based on the D&A residuals represent the 'true' observed variability, so should be distributed about the true value. In contrast, each control simulation reflects one possible realization of internal variability, which may be different from the 'real' sample. Thus, a difference in the medians between the two estimates does not necessarily reflect an inconsistency, but can arise from internal variability alone. However, the estimates from the observed situation should be within the confidence range of estimates from the control runs. This is indeed the case for 98.7% of estimates from the D&A residuals ($n_1 = 7,000$), which are within the 5%–95% range of control run estimates ($n_2 = 43$). Second, we calculate all possible comparisons (subtractions) between estimates from the D&A and control runs ($n = n_1 \times n_2 = 301,000$). To assess consistency of the lines of evidence, we tested whether the confidence range of these comparisons encompasses zero. Supplementary Fig. 8 (top) shows that it does, and in fact the value of zero is within the interquartile range of estimates.

The same observation is also valid for the right-hand side of Fig. 3b, indicating consistency between the estimates of unforced variability from reconstructions and simulations within the Medieval Quiet Period between 850–1100 CE (Supplementary Fig. 8, bottom). This is confirmed by a Kruskal–Wallis non-parametric test¹²² for differences of medians. Such non-parametric procedures may be sensitive to violation of the assumption of independence both within and between samples from populations. For the comparison of reconstructions and simulations for the Medieval Quiet Period, we expect that the largest non-independence stems from the definition of the period chosen. We therefore applied two different sensitivity tests: block-bootstrapping the period chosen, and bootstrapping the test statistic itself. Both sensitivity tests suggest that the inference is stable to sampling partly independent periods within the Medieval Quiet Period for the comparison.

Multidecadal trends. Fifty-one-year running linear GMST trends are calculated for each ensemble member of the interannual (unfiltered) reconstructions (Fig. 4). Trend values are allocated to the year at the end of each 51 yr period. The ensemble median across all 7,000 multimethod ensemble members (seven methods with a 1,000-member ensemble for each method) of each overlapping trend is displayed

as black line in Fig. 4a. The upper range of natural variability in these multidecadal trends is estimated using the 97.5th percentile across all trends and ensemble members between 1–1850 CE. For the model simulations, the corresponding 97.5th percentile is calculated back to the first year, where all simulations have data (850 CE) and for the control simulations across their respective period of data availability. Instrumental trends are calculated from the reconstruction target time series between 1850–2017 CE.

For Fig. 4b, the running linear trends are calculated for lengths between $n = 10$ and $n = 150$ yr for all 7,000 ensemble members. The fraction of ensemble members for which the centre year of the largest n -year trend over the Common Era falls into the post-1850 period is displayed. AR–noise proxy reconstructions shown in Fig. 4b were generated as follows: for each proxy record, a random noise time series with the same AR spectrum as the real proxy time series was generated using the `hosking.sim()` function in R. Each noise time series was truncated to the same length as the corresponding real proxy record. PCR, CPS and PAI reconstructions were then generated using these noise time series as input data, but the same parameters and target data as the real proxy reconstructions. Note that these noise proxies mimic the full AR spectrum of the real proxies, but the resulting noise proxies are uncorrelated with each other. Alternative results based on cross-correlated AR(1) noise proxies confirm our findings and are presented in Supplementary Fig. 20. The random numbers (brown line) are calculated by simply dividing the number of trend segments falling entirely within 1850–2000 (for example 140 for 10 yr trends) by the number of trends within the full 2,000 yr reconstruction period (1990 for 10 yr trends). This number decreases with increasing trend length.

EBM. The temporal evolution of the Earth's global energy balance (Fig. 4) was computed by solving the ordinary differential equation⁵⁷

$$C \frac{dT}{dt} = (1 - \alpha) \times Q - \sigma \epsilon T^4 \quad (7)$$

where C is the effective heat capacity of the system, set here to $c = 9.01 \times 10^8$ (in $\text{J m}^{-2} \text{K}^{-1}$) corresponding to an upper ocean box. The effective incoming radiation is reduced by the albedo ($\alpha = 0.3$) and depends on the incoming solar flux $Q = S_0/4 + \Delta F$, where S_0 is the solar constant ($S_0 = 1,365 \text{ W m}^{-2}$) and ΔF is a stochastic forcing term. The longwave outgoing radiation term is given by the product of the Stefan–Boltzmann constant $\sigma_B = 5.6710^{-8} \text{ W m}^{-2} \text{K}^{-4}$, and an emissivity term $\epsilon = 0.6$ with the fourth power of the system's temperature.

Here C is that of a 1 m^2 surface oceanic box, computed as $f \cdot \rho \cdot c \cdot h$, where $f = 0.7$ is the scaling for the approximate ocean area fraction of the planet, ρ the density of water, c the specific heat capacity of water and h the depth of the box. For a fixed value of Q , the ordinary differential equation solution describes the system's convergence to an equilibrium state. However, in a simplified experiment, volcanic perturbations can also be considered disturbances in the energy balance, modelled by a radiative forcing anomaly relative to the equilibrium state ΔF .

$$C \frac{dT}{dt} = (1 - \alpha) \times (Q + \Delta F) - \sigma \epsilon T^4 \quad (8)$$

We obtain ΔF (all = CO_2 + solar + volcanic, volcanic) from the forcing reconstructions of ref. ⁵³, translating original values to effective radiative forcing^{1,123} to units of W m^{-2} . Specifically, for the aerosol optical depth (AOD), we use -25 W m^{-2} per unit AOD¹. We run the EBM for two different forcing combinations: volcanic only (Fig. 4 and Supplementary Fig. 14) and all = CO_2 + solar + volcanic (Supplementary Fig. 14). The EBM response for both forcing scenarios shows consistent multidecadal variability (Supplementary Fig. 14). The EBM also demonstrates that with an increase of the frequency of volcanic events, the relative proportion of variance in the multidecadal to centennial band increases (Supplementary Fig. 15).

Data availability

The palaeotemperature records (PAGES 2k v2.0.0) used for all reconstructions are available at: www.ncdc.noaa.gov/paleo/study/21171. CMIP5 model runs are available at: <http://pcmdi9.llnl.gov/>. The primary outcomes for this study, including the temperature reconstructions for each method and the data used to construct the key figures including external forcing datasets used herein, model GMST and the screened input proxy data matrix, are available through the World Data Service (NOAA) Palaeoclimatology (<https://www.ncdc.noaa.gov/paleo/study/26872>) and Figshare (<https://doi.org/10.6084/m9.figshare.c.4507043>).

Code availability

The code to generate the figures is available along with the data in the repository listed above under Data availability.

References

72. Marcott, S. A., Shakun, J. D., Clark, P. U. & Mix, A. C. A reconstruction of regional and global temperature for the past 11,300 years. *Science* **339**, 1198–1201 (2013).

73. 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).
74. Mann, M. E., Rutherford, S., Wahl, E. & Ammann, C. Testing the fidelity of methods used in proxy-based reconstructions of past climate. *J. Clim.* **18**, 4097–4107 (2005).
75. Jones, P. et al. High-resolution palaeoclimatology of the last millennium: a review of current status and future prospects. *Holocene* **19**, 3–49 (2009).
76. Ljungqvist, F. C. A new reconstruction of temperature variability in the extra-tropical northern hemisphere during the last two millennia. *Geogr. Ann. Ser. A* **92**, 339–351 (2010).
77. Cook, E. R. et al. Asian monsoon failure and megadrought during the last millennium. *Science* **328**, 486–489 (2010).
78. Gergis, J., Neukom, R., Gallant, A. J. E. & Karoly, D. J. Australasian temperature reconstructions spanning the last millennium. *J. Clim.* **29**, 5365–5392 (2016).
79. Neukom, R. et al. Multiproxy summer and winter surface air temperature field reconstructions for southern South America covering the past centuries. *Clim. Dynam.* **37**, 35–51 (2011).
80. 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. *J. Clim.* **26**, 9194–9205 (2013).
81. Mann, M. E., Bradley, R. S. & Hughes, M. K. Global-scale temperature patterns and climate forcing over the past six centuries. *Nature* **392**, 779–787 (1998).
82. 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).
83. National Research Council *Surface Temperature Reconstructions for the Last 2,000 Years* (National Academies, 2006).
84. 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).
85. Wahl, E. R. & Ammann, C. M. Robustness of the Mann, Bradley, Hughes reconstruction of Northern Hemisphere surface temperatures: examination of criticisms based on the nature and processing of proxy climate evidence. *Climatic Change* **85**, 33–69 (2007).
86. McShane, B. B. & Wyner, A. J. A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Ann. Appl. Stat.* **5**, 5–44 (2011).
87. Wahl, E. R. & Smerdon, J. E. Comparative performance of paleoclimate field and index reconstructions derived from climate proxies and noise-only predictors. *Geophys. Res. Lett.* **39**, L06703 (2012).
88. Xoplaki, E. European spring and autumn temperature variability and change of extremes over the last half millennium. *Geophys. Res. Lett.* **32**, L15713 (2005).
89. Pauling, A., Luterbacher, J., Casty, C. & Wanner, H. Five hundred years of gridded high-resolution precipitation reconstructions over Europe and the connection to large-scale circulation. *Clim. Dynam.* **26**, 387–405 (2006).
90. Küttel, M. et al. The importance of ship log data: reconstructing North Atlantic, European and mediterranean sea level pressure fields back to 1750. *Clim. Dynam.* **34**, 1115–1128 (2010).
91. Neukom, R. et al. Multi-centennial summer and winter precipitation variability in southern South America. *Geophys. Res. Lett.* **37**, L14708 (2010).
92. Wang, J. et al. Internal and external forcing of multidecadal Atlantic climate variability over the past 1,200 years. *Nat. Geosci.* **10**, 512–517 (2017).
93. Schneider, T. Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. *J. Clim.* **14**, 853–871 (2001).
94. Fierro, R., Golub, G., Hansen, P. & O'Leary, D. Regularization by truncated total least squares. *SIAM J. Sci. Comput.* **18**, 1223–1241 (1997).
95. Shi, F., Yang, B. & Gunten, L. V. Preliminary multiproxy surface air temperature field reconstruction for China over the past millennium. *Sci. China Earth Sci.* **55**, 2058–2067 (2012).
96. Emile-Geay, J., Cobb, K. M., Mann, M. E. & Wittenberg, A. T. Estimating central equatorial pacific SST variability over the past millennium. Part I: methodology and validation. *J. Clim.* **26**, 2302–2328 (2013).
97. PAGES2k Consortium. Continental-scale temperature variability during the past two millennia. *Nat. Geosci.* **6**, 339–346 (2013).
98. Christiansen, B. & Ljungqvist, F. C. Reconstruction of the extratropical NH mean temperature over the last millennium with a method that preserves low-frequency variability. *J. Clim.* **24**, 6013–6034 (2011).
99. Tingley, M. P. & Huybers, P. A bayesian algorithm for reconstructing climate anomalies in space and time. Part II: comparison with the regularized expectation–maximization algorithm. *J. Clim.* **23**, 2782–2800 (2010).
100. Wang, Z. et al. Human-induced erosion has offset one-third of carbon emissions from land cover change. *Nat. Clim. Change* **7**, 345–349 (2017).
101. Blasone, R.-S. et al. Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov chain Monte Carlo sampling. *Adv. Water Resour.* **31**, 630–648 (2008).
102. Christiansen, B. & Ljungqvist, F. C. The extra-tropical Northern Hemisphere temperature in the last two millennia: reconstructions of low-frequency variability. *Clim. Past* **8**, 765–786 (2012).
103. Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M. & Karlén, W. Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data. *Nature* **433**, 613–617 (2005).
104. Shi, F. et al. Reconstruction of the Northern Hemisphere annual temperature change over the Common Era derived from tree rings. *Quat. Sci.* **35**, 1051–1063 (2015).
105. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B* **58**, 267–288 (1996).
106. Vieira, L. E. A., Solanki, S. K., Krivova, N. A. & Usoskin, I. Evolution of the solar irradiance during the Holocene. *Astron. Astrophys.* **531**, A6 (2011).
107. Toohey, M., Stevens, B., Schmidt, H. & Timmreck, C. Easy volcanic aerosol (EVA v1.0): an idealized forcing generator for climate simulations. *Geosci. Model Dev.* **9**, 4049–4070 (2016).
108. Meinshausen, M. et al. Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geosci. Model Dev.* **10**, 2057–2116 (2017).
109. Emile-Geay, J., Erb, M. P., Hakim, G. J., Steig, E. J. & Noone, D. C. Climate dynamics with the last millennium reanalysis. *PAGES Mag.* **25**, 162 (2017).
110. Steiger, N. J., Hakim, G. J., Steig, E. J., Battisti, D. S. & Roe, G. H. Assimilation of time-averaged pseudoproxies for climate reconstruction. *J. Clim.* **27**, 426–441 (2014).
111. Acevedo, W., Fallah, B., Reich, S. & Cubasch, U. Assimilation of pseudo-tree-ring-width observations into an atmospheric general circulation model. *Clim. Past* **13**, 545–557 (2017).
112. Landrum, L. et al. Last millennium climate and its variability in CCSM4. *J. Clim.* **26**, 1085–1111 (2013).
113. Dee, S. G., Steiger, N. J., Emile-Geay, J. & Hakim, G. J. On the utility of proxy system models for estimating climate states over the Common Era. *J. Adv. Model. Earth Syst.* **8**, 1164–1179 (2016).
114. Becker, A. et al. A description of the global land-surface precipitation data products of the global precipitation climatology centre with sample applications including centennial (trend) analysis from 1901–present. *Earth Syst. Sci. Data* **5**, 71–99 (2013).
115. Xiao-Ge, X., Tong-Wen, W. & Jie, Z. Introduction of CMIP5 experiments carried out with the climate system models of beijing climate center. *Adv. Clim. Change Res.* **4**, 41–49 (2013).
116. 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. *J. Adv. Model. Earth Syst.* **5**, 422–446 (2013).
117. Giorgetta, M. A. et al. Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the coupled model intercomparison project phase 5. *J. Adv. Model. Earth Syst.* **5**, 572–597 (2013).
118. Jungclaus, J. et al. *CMIP5 Simulations of the Max Planck Institute for Meteorology (MPI-M) based on the MPI-ESM-P model: The past1000 Experiment, Served by ESGF (WDCC at DKRZ, 2012); <https://doi.org/10.1594/WDCC/CMIP5.MXEPpk>*
119. 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).
120. Schmidt, G. A. et al. Present-day atmospheric simulations using GISS ModelE: comparison to in situ, satellite, and reanalysis data. *J. Clim.* **19**, 153–192 (2006).
121. Ribes, A., Planton, S. & Terray, L. Application of regularised optimal fingerprinting to attribution. Part I: method, properties and idealised analysis. *Clim. Dynam.* **41**, 2817–2836 (2013).
122. Kruskal, W. H. & Wallis, W. A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **47**, 583–621 (1952).
123. Myhre, G., Highwood, E. J., Shine, K. P. & Stordal, F. New estimates of radiative forcing due to well mixed greenhouse gases. *Geophys. Res. Lett.* **25**, 2715–2718 (1998).