



Swansea University
Prifysgol Abertawe



Cronfa - Swansea University Open Access Repository

This is an author produced version of a paper published in :

Environmental Research Letters

Cronfa URL for this paper:

<http://cronfa.swan.ac.uk/Record/cronfa26423>

Paper:

Luterbacher, J., Werner, J., Smerdon, J., Fernández-Donado, L., González-Rouco, F., Barriopedro, D., Ljungqvist, F., Büntgen, U., Zorita, E., Wagner, S., Esper, J., McCarroll, D., Toreti, A., Frank, D., Jungclaus, J., Barriendos, M., Bertolin, C., Bothe, O., Brázdil, R., Camuffo, D., Dobrovolný, P., Gagen, M., García-Bustamante, E., Ge, Q., Gómez-Navarro, J., Guiot, J., Hao, Z., Hegerl, G., Holmgren, K., Klimenko, V., Martín-Chivelet, J., Pfister, C., Roberts, N., Schindler, A., Schurer, A., Solomina, O., von Gunten, L., Wahl, E., Wanner, H., Wetter, O., Xoplaki, E., Yuan, N., Zanchettin, D., Zhang, H. & Zerefos, C. (2016). European summer temperatures since Roman times. *Environmental Research Letters*, 11(2), 024001

<http://dx.doi.org/10.1088/1748-9326/11/2/024001>

This article is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Authors are personally responsible for adhering to publisher restrictions or conditions. When uploading content they are required to comply with their publisher agreement and the SHERPA RoMEO database to judge whether or not it is copyright safe to add this version of the paper to this repository.

<http://www.swansea.ac.uk/iss/researchsupport/cronfa-support/>

Supplementary Online Material, SOM

European summer temperatures since Roman times

Euro-Med2k Consortium

Writing team: Luterbacher, J.*¹, Werner J.P.², Smerdon, J.E.³, Fernández-Donado, L.⁴, González-Rouco, F.J.⁴, Barriopedro, D.⁴, Ljungqvist, F.C.⁵, Büntgen, U.⁶, Zorita, E.⁷, Wagner, S.⁷, Esper J.⁸, McCarroll, D.⁹, Toreti, A.¹⁰, Frank, D.⁶, Jungclauss, J.H.¹¹

Barriendos¹², M., Bertolin, C.^{13,14}, Bothe, O.¹¹, Brázdil, R.¹⁵, Camuffo, D.¹³, Dobrovolný, P.¹⁵, Gagen, M.⁹, García-Bustamante, E.¹⁶, Ge, Q.¹⁷, Gómez-Navarro, J.J.¹⁸, Guiot, J.¹⁹, Hao, Z.¹⁷, Hegerl, G.C.²⁰, Holmgren, K.^{21,22}, Klimenko, V.V.²³, Martín-Chivelet, J.²⁴, Pfister, C.¹⁸, Roberts, N.²⁵, Schindler, A.²⁶, Schurer, A.²⁰, Solomina, O.²⁷, von Gunten, L.²⁸, Wahl, E.²⁹, Wanner, H.¹⁸, Wetter, O.¹⁸, Xoplaki, E.¹, Yuan, N.¹, Zanchettin, D.³⁰, Zhang, H.¹, and Zerefos, C.^{22,31}

* corresponding author

1 Department of Geography, Climatology, Climate Dynamics and Climate Change, Senckenbergstrasse 1, 35930 Giessen, Germany

2 University of Bergen, Department of Earth Science and Bjerknes Centre for Climate Research, Allégt. 41, 5020 Bergen, Norway

3 Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, 10964, USA

4 Instituto de Geociencias (IGEO), Centro Superior de Investigaciones Científicas, Universidad Complutense de Madrid, Ciudad Universitaria, 28040 Madrid, Spain

5 Department of History, Stockholm University, 106 91 Stockholm, Sweden and Bolin Centre for Climate Research, Stockholm University, 106 91 Stockholm, Sweden

6 Swiss Federal Research Institute WSL, 8903 Birmensdorf, Switzerland

7 Institute for Coastal Research, Helmholtz-Zentrum Geesthacht, 21502 Geesthacht, Germany

8 Department of Geography, Johannes Gutenberg University, 55099 Mainz, Germany

9 Department of Geography, Swansea University, Singleton Park, Swansea SA2 8PP, UK

10 European Commission, Joint Research Centre, 21027 Ispra, Italy

11 Max-Planck Institute for Meteorology, Bundesstrasse 53, 20146 Hamburg, Germany

12 Department of Modern History, University of Barcelona, Montalegre 6, 08001 Barcelona, Spain

13 National Research Council of Italy (CNR), Institute of Atmospheric Sciences and Climate (ISAC), Padova, Italy

14 University of Science and Technology NTNU -department of Architectural Design, History and Technology, research centre zero emission buildings, Trondheim, Norway

15 Institute of Geography, Masaryk University, and Global Change Research Centre AS CR, Brno, Czech Republic

16 Universidad de Murcia, Department of Physics, Murcia, Spain

17 Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, 11A, Datun Road, Chaoyang District, Beijing 100101, China

18 Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

- 19 Aix-Marseille Université, CNRS, IRD, CEREGE UM34, 13545 Aix en Provence, France
- 20 School of GeoSciences, University of Edinburgh, James Hutton Rd, Edinburgh EH9 3FE, UK
- 21 Department of Physical Geography, Stockholm University, SE-106 91 Stockholm, Sweden
- 22 Navarino Environmental Observatory, Costa Navarino, 24001, Messinia, Greece
- 23 Russian Academy of Science and Global Energy Problems Laboratory, Moscow Energy Institute, Krasnokazarmennaya St. 14, 111250 Moscow, Russia
- 24 Departamento Estratigrafía, Facultad de Ciencias Geológicas, Universidad Complutense de Madrid, 28040 Madrid, Spain
- 25 School of Geography, Earth and Environmental Sciences, Plymouth University, Plymouth PL4 8AA, UK
- 26 Federal Office of Meteorology and Climatology Meteoswiss, Operation Center 1, 8058 Zurich-Flughafen, Switzerland
- 27 Institute of Geography, Russian Academy of Science, Staromonetny-29, Moscow, Russia
- 28 PAGES International Project Office, Falkenplatz 16, 3012 Bern, Switzerland
- 29 NOAA Paleoclimatology Program, National Centers for Environmental Information, Center for Weather and Climate, Boulder, USA
- 30 University of Venice, Dept. of Environmental Sciences, Informatics and Statistics, 30123 Venice, Italy
- 31 Biomedical Research Foundation, Academy of Athens, Athens, Greece

Data

Proxy and instrumental data

We provide herein a short overview of the nine annually resolved tree-ring width (TRW), tree-ring maximum latewood density (MXD) and documentary records used for the reconstructions (Table S1).

The MXD-based summer temperature reconstruction from Northern Scandinavia (Nsc12, see Table S1) is derived from living and sub-fossil *Pinus sylvestris* trees from Finnish and Swedish Lapland. The record is longer and better replicated than any previously existing MXD-based temperature reconstruction, and has been used to assess inter-annual to millennial (Milankovitch) scale climate variability in Northern Scandinavia back to 138 BCE (Esper et al. 2012). Nsc12 integrates data from trees that fell into shallow lakes in Finnish Lapland with measurement series from living trees growing at the lake shores (to avoid ecological shifts that can limit the reliability of composite chronologies integrating relic material; see Tegel et al. 2010, D uthorn et al. 2013, 2015; Linderholm et al. 2014). The second tree-ring chronology spanning the past two millennia has been developed from records in the European Alps (Aus11, B untgen et al. 2011). It has been truncated in 138 BCE to enable comparison with the long-term record from Scandinavia. Additional records, that were recently developed and have not been integrated into any large-scale reconstructions include an updated MXD record from J amtland in central Sweden (Jae11, Gunnarson et al. 2011) extending back to 1107 CE, a TRW reconstruction from the French Alps (Fra12, B untgen et al. 2012) extending back to 969 CE and a composite record integrating MXD and TRW data from the Spanish Pyrenees back to 1260 CE (Pyr12, Dorado Li an et al. 2012b). These time series as well as the TRW record from Romania (Car09; Popa and Kern, 2009) and the MXD record from Switzerland (Swi06; B untgen et al. 2006) all extend into the 21st century and enable an assessment of proxy-based temperature variability until 2003 CE. All tree-ring chronologies used in this study have been detrended using the RCS technique (Esper et al. 2003) to avoid loss of low-frequency variance that can limit the assessment of temperature variations on centennial time scales (Cook et al. 1995; Esper et al. 2002). Table S1 provides information about the proxy data used.

Compared to the European mean summer reconstructions in PAGES 2k Consortium (2013) we excluded the TRW records from Slovakia (B untgen et al. 2013) and Albania (Seim et al. 2012), as they are not significantly correlated with the selected target of European summer temperature variability. Furthermore, the Tornetr ask MXD record of Briffa et al. (1992) that ends in 1980, has been exchanged with the updated and newly processed data by Melvin et al. (2013) and Esper et al. (2014).

Table S1: Proxy data information used for the European summer temperature reconstructions

Site	Country	Lon./Lat.	Elevation	Archive	Tree species	Proxy	Period	Reference
Tor13	Sweden	19.6° E/68.25° N	400 m	Tree-rings	Pine	MXD	500-2004	Melvin et al. (2013)
Jae11	Sweden	15° E/63.10° N	800-1000 m	Tree-rings	Pine	MXD	1107-2007	Gunnarson et al. (2011)
Nsc12	Finland	25° E/68° N	300 m	Tree-rings	Pine	MXD	138 BCE-2006	Esper et al. (2012)
Car09	Romania	25.3° E/47° N	1800 m	Tree-rings	Pine	TRW	1163-2005	Popa and Kern (2009)
Aus11	Austria	10.7° E/47° N	1450-2300 m	Tree-rings	Larch/Pine	TRW	500 BCE-2003	B�untgen et al. (2011)
Swi06	Switzerland	7.8° E/46.4° N	1600-2300 m	Tree-rings	Larch	MXD	755-2004	B�untgen et al. (2006)
Fra12	France	7.5° E/44° N	2100-2300 m	Tree-rings	Larch	TRW	969-2007	B�untgen et al. (2012)

Pyr12	Spain	1° E/42.5° N	1500-2500 m	Tree-rings	Pine	MXD/TRW	1260-2005	Dorado Liñán et al. (2012)
CEu10	Central Europe	45°-53° N, 6°-20° E	500-1500 m	Historical documents	–	–	1500-2007	Dobrovolný et al. (2010)

As in PAGES 2k Consortium (2013), we use the seasonally-resolved Central European temperature series (CEu10, Dobrovolný et al. 2010) that combines documentary and early instrumental data. A Central European index temperature series (the seven ordinally scaled ranks from -3 to 3) was created from documentary-based indices from Germany, Switzerland and the Czech Lands for the period 1500–1854 CE. The instrumental temperature series from 1760–2007 CE was taken as an average of eleven homogenized series: Kremsmünster, Vienna and Innsbruck (Austria), Basel, Geneva and Bern (Switzerland), Regensburg, Karlsruhe, Munich and Hohenpeissenberg (Germany) and Prague-Klementinum (the Czech Republic). These series were corrected for insufficient radiation protection (summer half-year) of early thermometers, and for the growth of the urban heat island effect before being used in the temperature reconstruction. Verification statistics indicated high reconstruction skill for all seasons (Dobrovolný et al. 2010).

Comparison between seasonal European mean temperatures using the raw data and the time and space continuous field infilled with the RegEM-Ridge algorithm

Figure S1 shows the differences between raw (with gaps) and infilled mean summer temperature indices derived from the CRUTEM4v temperature grid over the European domain. The differences are small and appear primarily prior to 1900 CE when the amount of missing data, particularly in the Mediterranean land region, is highest. The Pearson correlation coefficients between the proxy data and both European mean summer temperature and the local JJA grid cell temperatures in the infilled dataset were calculated over the 1850-2003 CE period and are shown in Table S2.

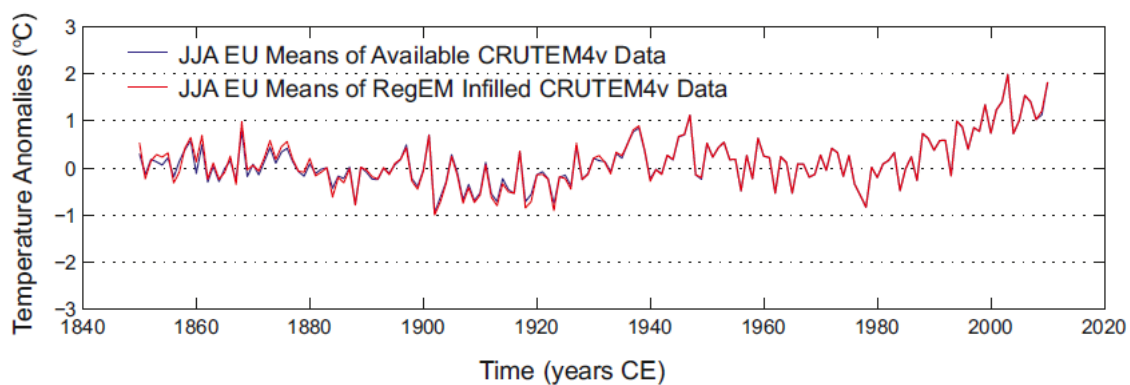


Figure S1: Comparison between the area-weighted JJA mean temperature anomalies (1961-1990 climatology) for the European domain using the raw (with gaps) data and the time and space continuous field infilled with the RegEM-Ridge algorithm (Schneider 2001).

Table S2: Start years, proxy abbreviation and Pearson's correlation coefficients (r) between the proxies and either the European regional mean summer temperature or the mean summer temperature of the grid containing the proxy (the latter number is given in parentheses) using the RegEM-Ridge

infilled temperature data. All correlations were computed for the 1850-2003 period. The null hypothesis of no correlation can be rejected at the 5% and 1% levels for correlations above 0.13 and 0.19, respectively (assuming a one-tailed PDF).

* includes instrumental data

Nest	Start Year (CE)	Proxy	Correlation (r)
1	-138	Nsc12	0.44 (0.76)
		Aus11	0.53 (0.69)
2	441	Tor13	0.51 (0.70)
3	755	Swi06	0.35 (0.61)
4	969	Fra12	0.45 (0.49)
5	1107	Jae11	0.47 (0.66)
6	1163	Car09	0.37 (0.41)
7	1260	Pyr10	0.13 (0.45)
8	1500	Ceu10	0.73 (0.96*)

Reconstruction Methods and Validation

Composite plus Scaling (CPS)

Multiple studies have previously applied the CPS method and validated its skill as a method for index reconstructions (e.g. Mann et al. 2008; PAGES 2k Consortium 2013). The initial calibration interval extended from 1850–1953 CE and was incremented by one year until reaching the final period of 1900–2003 CE, yielding a total of 51 reconstructions for each nest of Table S2. Within each calibration step, the 50 years excluded from calibration were used for validation. For each nest, the final CPS reconstruction was computed as the median reconstructed value in each year within the 51-member reconstruction ensemble.

Uncertainties were estimated based on the mean standard deviation (SD) of the residuals across all of the validation intervals by adding 1.96 times the SD estimate to the maximum and minimum ensemble values in each year, thereby providing the 95% confidence intervals of the median reconstruction. The final nested index reconstruction was combined by splicing together the median reconstruction and estimated uncertainties of each nest such that every reconstructed year is derived from the nest with the maximum number of possible predictors. Validation statistics across all reconstruction ensemble members within each nest indicate skilful CPS summer temperature reconstructions: The mean Reduction of Error (RE; Cook et al. 1994) and Coefficient of Efficiency (CE; Cook et al. 1994) cross validation statistics across the 51 intervals in each nest are positive (Table S3). Benchmarking experiments (Wahl and Smerdon 2012) with 1000 realizations of AR(1) red noise time series as predictors and autocorrelations approximating those of the available proxy records were also performed. These experiments yielded maximum mean RE and CE benchmark values below the mean values achieved for the actual reconstruction across all nests (Table S3). The mean Pearson's correlation coefficients for each nest were all above the 95% significance level ($r=0.245$) assuming a one-tailed test. Similar to the RE and CE results, the mean correlation coefficient across all validation intervals was also higher in each reconstruction nest than the maximum mean correlation achieved in the red noise benchmarking experiments (Table S3). Cross correlations between each of the derived nests during their period of overlap reflect strong coherence among the estimates (Table S4) and lend further support to the robustness of the reconstruction estimates of the individual nests.

Table S3: Mean validation statistics (in bold) for each nest of the CPS reconstruction. Numbers given in parentheses represent the mean, standard deviation and maximum mean validation statistics for

1000 benchmarking experiments performed for each nest using AR(1) red noise time series equal to the number of proxies in each nest and autocorrelations that approximate the estimated persistence of each proxy record.

Nest	Mean Correlation (r)	Mean RE	Mean CE
1	0.75 (0.01, 0.16, 0.52)	0.53 (-0.63, 0.32, 0.11)	0.40 (-1.15, 0.44, -0.15)
2	0.74 (-0.01, 0.16, 0.47)	0.47 (-0.63, 0.31, 0.12)	0.31 (-1.15, 0.41, -0.15)
3	0.76 (0.02, 0.16, 0.54)	0.47 (-0.59, 0.31, 0.29)	0.32 (-1.10, 0.42, 0.06)
4	0.78 (0.00, 0.17, 0.55)	0.50 (-0.60, 0.33, 0.27)	0.36 (-1.12, 0.44, 0.04)
5	0.79 (0.00, 0.17, 0.55)	0.44 (-0.60, 0.33, 0.23)	0.27 (-1.11, 0.44, -0.01)
6	0.80 (0.00, 0.17, 0.48)	0.50 (-0.58, 0.34, 0.25)	0.35 (-1.10, 0.46, 0.03)
7	0.80 (0.01, 0.17, 0.55)	0.50 (-0.57, 0.32, 0.34)	0.35 (-1.07, 0.42, 0.13)
8	0.85 (0.02, 0.16, 0.48)	0.63 (-0.53, 0.30, 0.18)	0.52 (-1.03, 0.40, -0.09)

Table S4: Pearson's correlation coefficients between each CPS nest during the 1500-1849 CE common interval.

	Nest1	Nest2	Nest3	Nest4	Nest5	Nest6	Nest7	Nest8
Nest1	1	0.91	0.89	0.85	0.84	0.85	0.85	0.74
Nest2		1	0.96	0.93	0.92	0.93	0.92	0.76
Nest3			1	0.98	0.97	0.97	0.97	0.85
Nest4				1	0.97	0.97	0.97	0.85
Nest5					1	0.99	0.99	0.87
Nest6						1	0.99	0.88
Nest7							1	0.89
Nest8								1

Figure S2 shows a comparison between the new CPS based reconstruction (without two tree ring predictors including the updated Torneträsk MXD record and the one published in PAGES 2k Consortium (2013)). The two reconstructions are virtually identical and have a Pearson's correlation coefficient of 0.99 over the full reconstruction period.

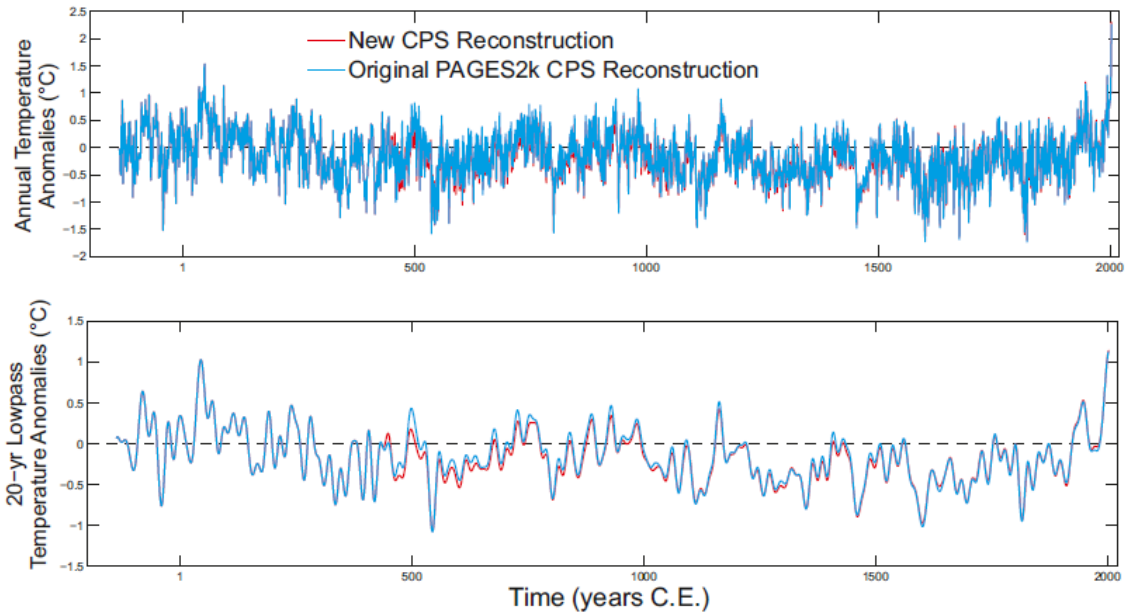


Figure S2: Comparison between the new CPS based European summer mean reconstruction and the one published in PAGES 2k Consortium (2013).
Bottom: as in the top panel but for 20-year low-pass filtered time-series.

The limited number of proxies might be an important caveat for the reconstructions. Figure S3 presents the sensitivity of the reconstruction to the number of underlying proxies. The descriptions of the nests and their quantitative comparisons are given in Tables S2-S4. While there are small differences between some nests, the covariance among each nest is remarkably consistent across all of the nests during their periods of overlap.

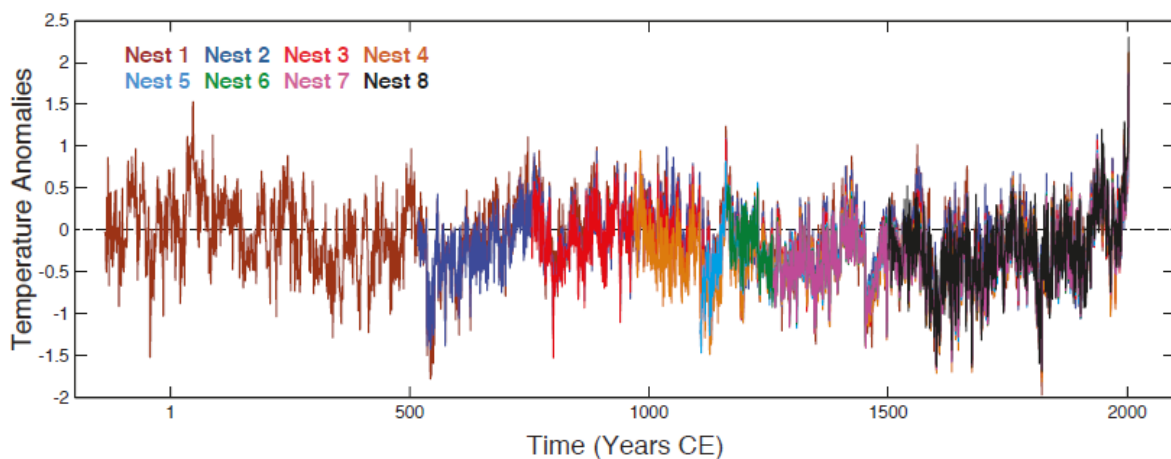


Figure S3: Comparison of the 8 nests in the CPS-based reconstructions of mean European summer temperature anomalies for the period 138 BCE-2003 CE. The descriptions of the nests and their quantitative comparisons are given in Tables S2-S4.

In addition to the CPS methodology, alternative techniques were explored for application to the longest nest in the European mean summer temperature reconstruction (not shown). Ordinary least squares (OLS) regression was employed by using the same resampling scheme discussed in the methods section of the article. OLS was also tested with a static validation/calibration interval of 1850-1899 /1900-2003 CE and the residual modelling and

Monte Carlo resampling described in Wahl and Smerdon (2012). While both iterations of OLS were similar to the corresponding CPS scenarios, the validation statistics for the OLS reconstructions were overall poorer than for CPS. Additionally, OLS reduced the variability more than CPS, although the difference of the validation period mean temperature was reconstructed too cold in the static-period OLS exercises. The RegEM-EIV method (Mann et al. 2008) was also tested using the same static validation/calibration intervals of 1850-1899 / 1900-2003 CE (not shown). This approach returned similar results as the CPS or the OLS methods in corresponding scenarios. Despite the above-described differences in validation statistics, the general characteristics of all derived reconstructions were comparable and provide further validation of the CPS nested temperature reconstruction.

Bayesian Hierarchical Modelling (BHM)

For the summer temperature reconstruction at grid point scale, we used a slightly modified version of the simple hierarchical model implemented by Werner et al. (2013) and Tingley and Huybers (2013). Here, we provide details on temperature field \mathbf{T}_t to be reconstructed, the field of the instrumental records \mathbf{I}_t , as well as the field of the temperature proxies \mathbf{P}_t . The assumptions on the inter-annual variability of each field and their interdependencies are subsequently given. Proxy data are centred to have zero mean during the 1500–2000 CE period to facilitate data processing. In contrast to the CPS application, the data-infilling step for the instrumental temperature observations was not necessary.

The hierarchical model in a paleoclimate reconstruction consists of a process-level model, describing the spatio-temporal evolution of the climate field, and a data-level that describes how proxies and instrumental data record the climate anomalies. The temperatures drive the realizations of the instrumental observations and the proxies through linear stochastic response functions. From these model assumptions, the corresponding conditional probability densities for both the temperature field and the model parameters can be derived. These are then combined with prior probability densities reflecting knowledge on the system using Bayes' rule. Here, the prior distributions are left as being very broad and uninformative (see parameters of Tingley and Huybers 2010a; Werner et al. 2013). The resulting posterior probability densities are estimated by using a Gibbs sampler. In contrast to Werner et al. (2013) and Tingley and Huybers (2013), in which all proxies of one type were constructed equally and thus shared a common proxy response function with one set of parameters, the reconstructions derived herein use a distinct parameter for each single proxy.

The equations as used also by Tingley and Huybers (2013) are:

$$\mathbf{T}_{t+1} - \mu = \alpha(\mathbf{T}_t - \mu) + \boldsymbol{\varepsilon}_{T,t} \quad (1a)$$

$$\mathbf{I}_t = \mathbf{H}_t(\mathbf{T}_t + \boldsymbol{\varepsilon}_{I,t})^T \quad (1b)$$

$$\mathbf{P}_t = \mathbf{H}_t(\beta_0 + \beta_1 \mathbf{T}_t + \beta_2 \mathbf{P}_{t-1} + \boldsymbol{\varepsilon}_{P,t})^T \quad (1c)$$

$$\boldsymbol{\varepsilon}_{T,t} \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \quad \Sigma_{ij} = \sigma^2 \exp(-\varphi |x_i - x_j|) \quad (1d)$$

$$\boldsymbol{\varepsilon}_{I,t} \sim N(\mathbf{0}, \mathbf{I} \tau_I^2) \quad (1e)$$

$$\boldsymbol{\varepsilon}_{P,t} \sim N(\mathbf{0}, \mathbf{I} \tau_P^2) \quad (1f)$$

Equation (1a) describes the process-level model, the temporal evolution of the temperature field \mathbf{T}_t for time index t , using an AR(1) process with uniform persistence α and mean μ . The inter-annual variability of the temperature field $\boldsymbol{\varepsilon}_{T,t}$ is multivariate normal with zero mean and covariance matrix $\boldsymbol{\Sigma}$, eq. (1d). The spatial covariance matrix is homogeneous and decreases exponentially with orthodromic distance with a spatial correlation length of φ . The variance of the inter-annual local fluctuation is given by σ^2 . The instrumental observations \mathbf{I}_t are assumed

to be noisy representations of the true temperature at the location of observation (eq. 1b) with noise strength (variance) τ_l^2 (eq. 1e). The natural climate archives P_t are modelled with a linear stochastic response function (eq. 1c) for latewood density and documentary evidence. The tree-ring width of one year is influenced, however, by the previous year's growth (Frank et al. 2007), which might be responsible for the additional long term memory as observed by Zhang et al (2015). The term β_2 was therefore included to estimate this influence for TRW data. It was held fixed at zero for the other proxy types. The proxy noise strength is τ_p^2 . While the model does allow for one proxy type to be used at more than one location we have chosen to estimate all coefficients for each proxy series. This is in contrast to the original articles by Tingley and Huybers (2010a,b) and also Tingley and Huybers (2013), in which a single proxy transfer function with only one set of global parameters was used for more than one location. The H_t matrices contain ones at locations where proxy or instrumental data are present at time step t , and zero otherwise.

As proposed by Gelman et al. (2003), the reconstruction using Bayesian inference is done in two steps. The first step is the actual inference step, performed with a Gibbs sampler over a subset of the proxy data and the full instrumental data. The second step consist of a predictive experiment following Gelman et al. (2003); the parameters are sampled from the results in the inference step and temperatures are predicted while withholding the instrumental data, resulting in the full, two millennia long climate field reconstruction.

For the inference step, input data are first truncated to start at 755 CE, when four proxy series are available, and end at 2003 CE. The truncation helps avoid convergence issues prior to 754 CE when only two proxies are available. The Bayesian inversion is done using five chains of the Gibbs sampler that are run for 5000 steps each. The last half is checked for convergence using the Potential Scale Reduction Factor *Rhat* measure as defined in Gelman et al. (2003). All four chains converge after about 2000 steps and the *Rhat* values do not deviate more than 10^{-3} from unity. These Gibbs sampler runs are used to create a set of likely model parameter distributions. These are then used to perform predictive experiments for the temperature field over the full reconstruction period 138 BCE to 2003 CE using the complete proxy data while withholding the instrumental observations. The predictive experiment is run as a (simplified) Gibbs sampler, initialized with estimates from the first run. The parameters are sampled from the distributions estimated in the inference step and the temperatures are estimated conditional on these. In this way the parameter estimation is not influenced by a data set that is too sparse in the past (before 755 CE). Additionally, the reconstructions over the instrumental period can be checked against the instrumental dataset. Again, only the last iterations of the Gibbs sampler are used, this time as the final ensemble of gridded temperature reconstructions. From this ensemble we calculate measures, usually showing the ensemble mean of the result as well as the upper and lower 2.5% quantiles of the distribution as uncertainty estimates.

To assess whether a uniform persistence term α is appropriate in this context, we apply the Kramers-Moyal-Expansion (KME, Risken 1989, Stemler et al. 2007, Werner et al. 2014) to the instrumental data on the grid scale. From the drift term in the stochastic function we can derive the persistence term α . The results are shown in Fig. S3 (left panel). The persistence is roughly uniform over the landmass. The local strength of the inter-annual variability, σ , of summer temperatures is presented in Fig. S4 (centre panel). Local inter-annual variability is strongest in the Scandinavian/western Russian regions. Deviations from the additive white noise model (eq. 1a) are shown in Fig. S4 (right panel). The results support our approach of first rescaling the instrumental data to zero mean and uniform standard deviation and then assuming homogenous model parameters in the reconstruction, but note that the amount of data used in the KME is insufficient to obtain exact values for the second and higher-order terms.

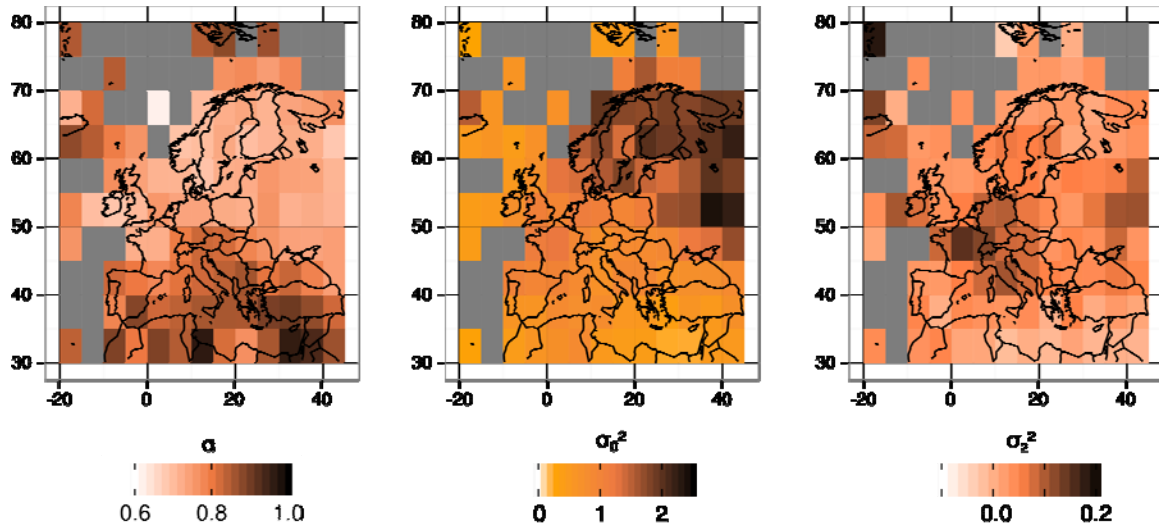


Figure S4: Persistence, inter-annual variability and deviation from additive white noise model (left, centre and right panel respectively), estimated through Kramers-Moyal-Expansion from instrumental data CRUTEM 4v (Jones et al. 2012), JJA, 1850-2010. Persistence is close to 0.7 for most of the grid cells. The difference in variability is removed by standardization of data prior to the reconstruction. Deviations from the additive white noise model are small.

Histograms of the proxy Signal to Noise Ratio (SNR) calculated from the drawn parameters are shown in Fig. S5. Most of the tree-ring data show high SNR values, above and exceeding 1, with the exception of the data from the Pyrenees. Of the natural climate archives, the MXD tree-ring series from Torneträsk (Melvin et al. 2013) and Northern Scandinavia (Esper et al. 2012) show the strongest values. The SNR of the documentary data is artificially inflated, as it contains instrumental data in the calibration period.

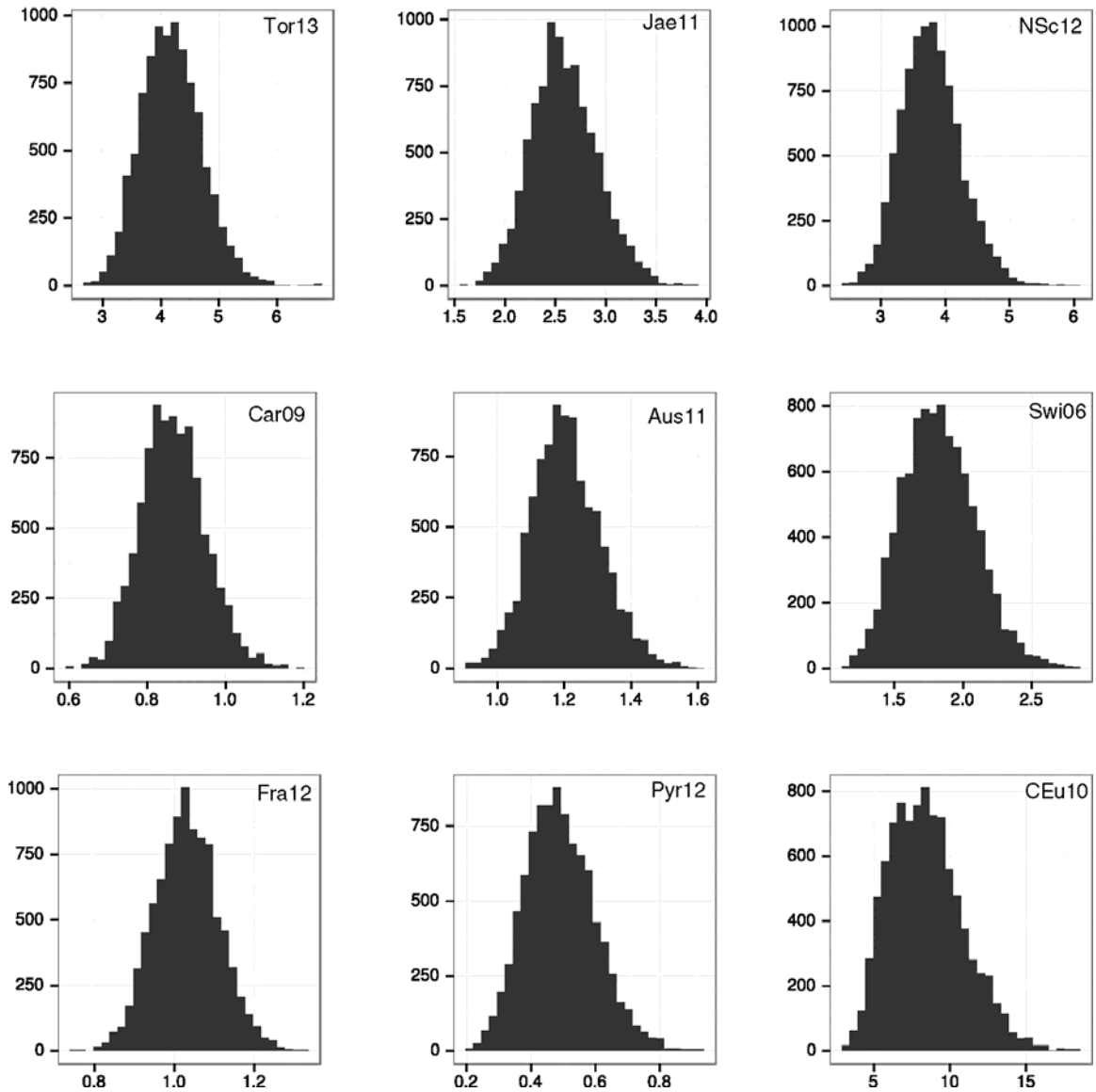


Figure S5: Histograms of estimated SNR for each proxy calculated from the parameters inferred from the BHM. Details on the proxy data can be found in Table S1.

Figure S6 shows draws of the parameters for the spatiotemporal temperature model (eq. 1a), the persistence α , the mean μ , the strength of the inter-annual variability σ , and the spatial correlation length φ , as well as the instrumental noise variance τ_I^2 . The model parameters are close to the results of the preliminary analysis in Fig. S4.

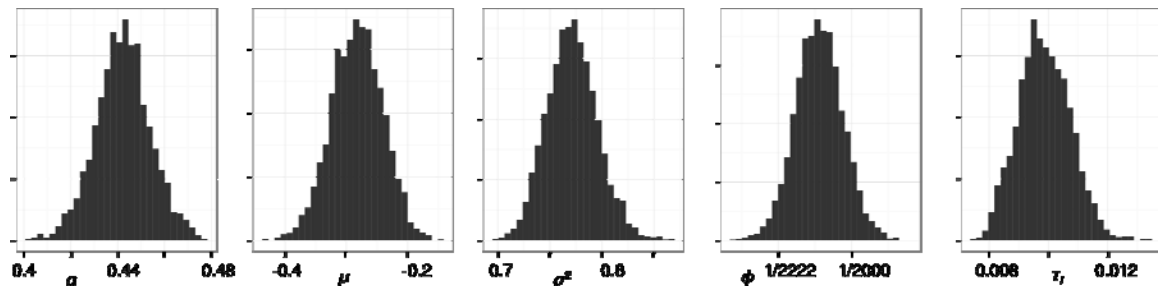


Figure S6: Histograms of the model parameters (first four panels, eq. 1a) and the strength of the instrumental noise (right panel, eq. 1b).

Using all of the predictors used in the European reconstruction of PAGES 2k Consortium (2013) results in zero signal content for the two here excluded (Tatra and Albanian) tree ring series. Thus, the resulting reconstruction is virtually identical – that is, the posterior distributions of the temperatures at all locations cannot be distinguished. The BHM algorithm was therefore able to filter out a posteriori unsuitable set of proxy data.

Comparison of recent and earlier warm periods

To test whether recent warm conditions are unusual in the context of the two full reconstructions presented in Fig. 1C we made a series of simple comparisons using 2-sample t tests. We present results for individual t tests, but also for tests that have been corrected for first order autocorrelation by adjusting the variance and estimating the equivalent sample size and thus the adjusted degrees of freedom. To take account of the increased likelihood of ‘false positive’ results that arise with multiple testing, we use two methods: a false discovery rate controlling method (FDR; Benjamini and Hochberg, 1995) and the much harsher Bonferroni correction (Dunn, 1961; Field, 2013). We use two-tail tests because we did not predict the sign of the differences in advance.

Comparing the mean temperature anomaly of the 20th century with every other century the two reconstructions differ only slightly. In the CPS reconstruction only the 1st century CE has a mean temperature anomaly higher than that of the 20th century and the difference is statistically significant at the 5% level using the simple t test and remains so when first order autocorrelation is taken into account (Table S5). However, after correction for multiple testing the difference is not significant. In the BHM reconstruction (Table S6) there are two century means that are very slightly (<0.1 °C) warmer but neither is significantly different from the 20th century even using a simple t test with no corrections. The pattern of cold centuries is similar but not identical in the two reconstructions. Using the CPS data 14 centuries are significantly different from (colder than) the 20th using the FDR correction and 12 remain so even using the Bonferroni correction. Using the BHM data the equivalent values are 14 and 11. In both reconstructions the first four centuries of the first millennium and the 8th and 10th centuries are not significantly different from the 20th. The centuries within the Little Ice Age are strongly significantly different from the 20th century in both reconstructions.

The same procedure was repeated using the second half of the 20th century as the reference period and comparing with every other half century. In the CPS reconstruction (Table S7) there are four half centuries, all in the early first millennium, that are very slightly warmer but the differences are not statistically significant even using simple t tests with no corrections. Using the BHM data (Table S8), there are three half centuries that are slightly warmer but again the differences are not significant even using simple t tests with no corrections. The pattern of significant cold half centuries largely mirrors the pattern shown for warm centuries.

The last 30 years in both reconstructions are very warm relative to the full records and when compared to contiguous sets of 30 years (Tables S9 and S10) there is only one interval (24 to 53CE) that is slightly warmer. These samples are too short to apply a correction for autocorrelation, but after applying a simple t test the warmest 30-year interval is not significantly different from the last 30 years of the proxy records.

Although the proxy-based climate reconstructions only extend as far as 2003, the equivalent instrumental data used for calibration and scaling extends to cover the summer of 2015. The mean temperature of the past 30 years in the instrumental data (1986-2015) is much higher than that of any 30-year period over the last two millennia. When compared with the same contiguous 30-year periods used above (Tables S11 and S12) there is only one, in the 1st century (24-53 CE), that is not significantly colder even when using the harshest correction for family-wise error. It is extremely likely that the average European summer temperatures of the last 30 years have been anomalously warm with respect to the last two millennia.

Table S5. Mean temperature anomaly for each century and the difference (Diff) from the 20th century using the CPS reconstruction. Significance levels are given for each individual two-tail t-test, for tests taking account of first order autocorrelation (Cor AC) and then using two corrections for multiple testing; the False Discovery Rate ($\alpha = 0.05$) and the Bonferroni correction for family-wise error ($\alpha = 0.05$). Warmer periods and significant differences are identified in orange, significant cooler periods in blue ($p < 0.05$).

CPS data: centuries						
C	Mean	Diff	T test	Cor AC	FDR	Bonf
20	0.09					
19	-0.32	-0.41	<0.001	<0.001	Yes	Yes
18	-0.24	-0.34	<0.001	<0.001	Yes	Yes
17	-0.52	-0.61	<0.001	<0.001	Yes	Yes
16	-0.29	-0.38	<0.001	<0.001	Yes	Yes
15	-0.26	-0.35	<0.001	<0.001	Yes	Yes
14	-0.41	-0.50	<0.001	<0.001	Yes	Yes
13	-0.33	-0.42	<0.001	<0.001	Yes	Yes
12	-0.24	-0.33	<0.001	<0.001	Yes	Yes
11	-0.26	-0.35	<0.001	<0.001	Yes	Yes
10	0.08	-0.01	0.823	0.850	No	No
9	-0.10	-0.20	0.001	0.006	Yes	No
8	0.09	-0.01	0.922	0.936	No	No
7	-0.18	-0.27	<0.001	<0.001	Yes	Yes
6	-0.26	-0.36	<0.001	<0.001	Yes	Yes
5	-0.13	-0.22	<0.001	0.005	Yes	No
4	-0.37	-0.46	<0.001	<0.001	Yes	Yes
3	0.02	-0.07	0.245	0.340	No	No
2	0.04	-0.06	0.340	0.439	No	No
1	0.27	0.17	0.011	0.043	No	No
0	0.06	-0.03	0.635	0.708	No	No

Table S6. As Table S5 but for the BHM reconstruction

BHM data: centuries						
C	Mean	Diff	T test	Cor AC	FDR	Bonf
20	0.16					
19	-0.32	-0.48	<0.001	<0.001	Yes	Yes
18	-0.25	-0.42	<0.001	<0.001	Yes	Yes
17	-0.51	-0.68	<0.001	<0.001	Yes	Yes
16	-0.31	-0.47	<0.001	<0.001	Yes	Yes
15	-0.24	-0.40	<0.001	<0.001	Yes	Yes
14	-0.32	-0.48	<0.001	<0.001	Yes	Yes
13	-0.07	-0.23	0.001	0.010	Yes	No
12	0.00	-0.17	0.036	0.101	No	No
11	-0.20	-0.36	<0.001	<0.001	Yes	Yes
10	0.25	0.08	0.227	0.316	No	No
9	-0.02	-0.19	0.005	0.022	Yes	No
8	0.04	-0.13	0.051	0.102	No	No
7	-0.21	-0.37	<0.001	<0.001	Yes	Yes
6	-0.33	-0.49	<0.001	<0.001	Yes	Yes
5	-0.12	-0.28	<0.001	0.001	Yes	Yes
4	-0.34	-0.50	<0.001	<0.001	Yes	Yes
3	-0.03	-0.19	0.003	0.015	Yes	No
2	0.02	-0.14	0.022	0.065	No	No
1	0.26	0.09	0.185	0.299	No	No
0	0.07	-0.09	0.195	0.311	No	No

Table S7. As Table S5 but for the mean temperature anomaly of each half century and the difference (Diff) from the second half of the 20th century using the CPS reconstruction

CPS data: half centuries						
C	Mean	Diff	T test	Cor AC	FDR	Bonf
20	0.18					
20	0.01	-0.17	0.072	0.151	No	No
19	-0.19	-0.37	<0.001	<0.001	Yes	Yes
19	-0.44	-0.62	<0.001	<0.001	Yes	Yes
18	-0.14	-0.31	<0.001	0.001	Yes	Yes
18	-0.35	-0.52	<0.001	<0.001	Yes	Yes
17	-0.47	-0.65	<0.001	<0.001	Yes	Yes
17	-0.56	-0.74	<0.001	<0.001	Yes	Yes
16	-0.40	-0.58	<0.001	<0.001	Yes	Yes
16	-0.17	-0.35	<0.001	<0.001	Yes	Yes
15	-0.50	-0.68	<0.001	<0.001	Yes	Yes
15	-0.02	-0.19	0.007	0.010	Yes	No
14	-0.34	-0.52	<0.001	<0.001	Yes	Yes
14	-0.47	-0.65	<0.001	<0.001	Yes	Yes
13	-0.48	-0.66	<0.001	<0.001	Yes	Yes
13	-0.18	-0.36	<0.001	<0.001	Yes	Yes
12	0.03	-0.15	0.063	0.129	No	No
12	-0.50	-0.68	<0.001	<0.001	Yes	Yes
11	-0.25	-0.42	<0.001	<0.001	Yes	Yes
11	-0.27	-0.45	<0.001	<0.001	Yes	Yes
10	0.11	-0.07	0.379	0.440	No	No
10	0.05	-0.13	0.142	0.243	No	No
9	-0.02	-0.19	0.024	0.069	No	No
9	-0.19	-0.37	<0.001	<0.001	Yes	Yes
8	0.03	-0.14	0.110	0.205	No	No
8	0.14	-0.04	0.638	0.687	No	No
7	-0.11	-0.29	<0.001	0.002	Yes	No
7	-0.24	-0.42	<0.001	<0.001	Yes	Yes
6	-0.28	-0.46	<0.001	<0.001	Yes	Yes
6	-0.25	-0.42	<0.001	0.002	Yes	No
5	-0.10	-0.28	0.002	0.012	Yes	No
5	-0.16	-0.34	<0.001	0.004	Yes	No
4	-0.33	-0.50	<0.001	<0.001	Yes	Yes
4	-0.41	-0.58	<0.001	<0.001	Yes	Yes
3	-0.04	-0.22	0.011	0.042	No	No
3	0.08	-0.09	0.266	0.379	No	No
2	-0.11	-0.29	<0.001	0.006	Yes	No
2	0.19	0.01	0.869	0.878	No	No
1	0.25	0.07	0.397	0.491	No	No
1	0.28	0.11	0.288	0.419	No	No
0	-0.10	-0.28	0.004	0.026	Yes	No

0	0.22	0.04	0.607	0.694	No	No
---	------	------	-------	-------	----	----

Table S8. As Table S7 but for the BHM reconstruction

BHM data: half centuries						
C	Mean	Diff	T test	Cor AC	FDR	Bonf.
20	0.22					
20	0.11	-0.11	0.326	0.432	No	No
19	-0.16	-0.38	<0.001	<0.001	Yes	Yes
19	-0.47	-0.69	<0.001	<0.001	Yes	Yes
18	-0.17	-0.39	<0.001	<0.001	Yes	Yes
18	-0.34	-0.55	<0.001	<0.001	Yes	Yes
17	-0.47	-0.69	<0.001	<0.001	Yes	Yes
17	-0.56	-0.77	<0.001	<0.001	Yes	Yes
16	-0.43	-0.65	<0.001	<0.001	Yes	Yes
16	-0.19	-0.41	<0.001	<0.001	Yes	Yes
15	-0.46	-0.68	<0.001	<0.001	Yes	Yes
15	-0.01	-0.23	0.003	0.003	Yes	No
14	-0.25	-0.47	<0.001	<0.001	Yes	Yes
14	-0.38	-0.60	<0.001	<0.001	Yes	Yes
13	-0.34	-0.55	<0.001	<0.001	Yes	Yes
13	0.20	-0.02	0.801	0.846	No	No
12	0.28	0.06	0.506	0.597	No	No
12	-0.28	-0.50	<0.001	<0.001	Yes	Yes
11	-0.21	-0.43	<0.001	0.001	Yes	Yes
11	-0.19	-0.40	<0.001	<0.001	Yes	Yes
10	0.30	0.08	0.363	0.440	No	No
10	0.20	-0.02	0.813	0.851	No	No
9	0.02	-0.19	0.042	0.106	No	No
9	-0.07	-0.29	<0.001	0.003	Yes	No
8	-0.01	-0.23	0.013	0.046	No	No
8	0.09	-0.13	0.089	0.142	No	No
7	-0.14	-0.36	<0.001	<0.001	Yes	Yes
7	-0.28	-0.50	<0.001	<0.001	Yes	Yes
6	-0.38	-0.60	<0.001	<0.001	Yes	Yes
6	-0.27	-0.49	<0.001	0.001	Yes	Yes
5	-0.10	-0.31	0.001	0.006	Yes	No
5	-0.14	-0.36	<0.001	0.003	Yes	No
4	-0.27	-0.48	<0.001	<0.001	Yes	Yes
4	-0.41	-0.63	<0.001	<0.001	Yes	Yes
3	-0.08	-0.30	0.001	0.004	Yes	No
3	0.03	-0.19	0.026	0.073	No	No
2	-0.12	-0.34	<0.001	0.002	Yes	No
2	0.16	-0.06	0.464	0.523	No	No

1	0.21	-0.003	0.969	0.974	No	No
1	0.30	0.08	0.426	0.547	No	No
0	0.00	-0.21	0.026	0.086	No	No
0	0.15	-0.07	0.408	0.523	No	No

Table S9. As Table S5 but for the mean temperature anomaly of contiguous 30-year periods and the difference (Diff) from the last 30 years of the record using the CPS reconstruction.

CPS data: 30 years						
start	End	Mean	Diff	t test	FDR	Bonf.
1974	2003	0.39				
1944	1973	0.19	-0.20	0.156	No	No
1914	1943	0.07	-0.32	0.019	Yes	No
1884	1913	-0.34	-0.73	<0.001	Yes	Yes
1854	1883	-0.13	-0.51	<0.001	Yes	Yes
1824	1853	-0.33	-0.71	<0.001	Yes	Yes
1794	1823	-0.46	-0.85	<0.001	Yes	Yes
1764	1793	-0.22	-0.60	<0.001	Yes	Yes
1734	1763	-0.17	-0.55	<0.001	Yes	Yes
1704	1733	-0.36	-0.74	<0.001	Yes	Yes
1674	1703	-0.59	-0.98	<0.001	Yes	Yes
1644	1673	-0.31	-0.70	<0.001	Yes	Yes
1614	1643	-0.49	-0.87	<0.001	Yes	Yes
1584	1613	-0.79	-1.18	<0.001	Yes	Yes
1554	1583	-0.24	-0.63	<0.001	Yes	Yes
1524	1553	-0.19	-0.58	<0.001	Yes	Yes
1494	1523	-0.18	-0.56	<0.001	Yes	Yes
1464	1493	-0.40	-0.78	<0.001	Yes	Yes
1434	1463	-0.41	-0.79	<0.001	Yes	Yes
1404	1433	0.01	-0.37	0.004	Yes	No
1374	1403	-0.29	-0.68	<0.001	Yes	Yes
1344	1373	-0.47	-0.85	<0.001	Yes	Yes
1314	1343	-0.36	-0.75	<0.001	Yes	Yes
1284	1313	-0.47	-0.86	<0.001	Yes	Yes
1254	1283	-0.50	-0.88	<0.001	Yes	Yes
1224	1253	-0.34	-0.72	<0.001	Yes	Yes
1194	1223	-0.03	-0.42	0.001	Yes	No
1164	1193	-0.03	-0.42	0.002	Yes	No
1134	1163	-0.08	-0.47	0.001	Yes	No
1104	1133	-0.64	-1.02	<0.001	Yes	Yes
1074	1103	-0.24	-0.63	<0.001	Yes	Yes
1044	1073	-0.26	-0.65	<0.001	Yes	Yes
1014	1043	-0.29	-0.68	<0.001	Yes	Yes
984	1013	-0.03	-0.41	0.002	Yes	No

954	983	0.11	-0.27	0.042	No	No
924	953	0.18	-0.21	0.135	No	No
894	923	-0.08	-0.47	0.001	Yes	Yes
864	893	0.14	-0.25	0.069	No	No
834	863	-0.13	-0.52	<0.001	Yes	Yes
804	833	-0.28	-0.67	<0.001	Yes	Yes
774	803	-0.21	-0.60	<0.001	Yes	Yes
744	773	0.26	-0.13	0.330	No	No
714	743	0.19	-0.20	0.118	No	No
684	713	-0.11	-0.50	<0.001	Yes	Yes
654	683	-0.06	-0.45	0.001	Yes	No
624	653	-0.25	-0.63	<0.001	Yes	Yes
594	623	-0.26	-0.65	<0.001	Yes	Yes
564	593	-0.29	-0.68	<0.001	Yes	Yes
534	563	-0.55	-0.93	<0.001	Yes	Yes
504	533	0.00	-0.38	0.004	Yes	No
474	503	0.06	-0.33	0.022	Yes	No
444	473	-0.20	-0.58	<0.001	Yes	Yes
414	443	-0.27	-0.65	<0.001	Yes	Yes
384	413	-0.22	-0.60	<0.001	Yes	Yes
354	383	-0.25	-0.64	<0.001	Yes	Yes
324	353	-0.46	-0.85	<0.001	Yes	Yes
294	323	-0.28	-0.66	<0.001	Yes	Yes
264	293	-0.06	-0.45	0.002	Yes	No
234	263	0.23	-0.15	0.259	No	No
204	233	-0.05	-0.44	0.002	Yes	No
174	203	0.01	-0.38	0.005	Yes	No
144	173	-0.26	-0.65	<0.001	Yes	Yes
114	143	0.26	-0.13	0.293	No	No
84	113	0.15	-0.24	0.074	No	No
54	83	0.29	-0.10	0.455	No	No
24	53	0.55	0.17	0.274	No	No
-7	23	0.08	-0.30	0.030	Yes	No
-37	-8	-0.05	-0.44	0.002	Yes	No
-67	-38	-0.08	-0.47	0.003	Yes	No
-97	-68	0.29	-0.10	0.460	No	No
-127	-98	-0.10	-0.49	0.001	Yes	Yes

Table S10. As Table S9 but for the BHM reconstruction.

	BHM data: 30 years					
start	End	Mean	Diff	t test	FDR	Bonf
1974	2003	0.42				
1944	1973	0.27	-0.15	0.334	No	No
1914	1943	0.20	-0.22	0.129	No	No
1884	1913	-0.34	-0.76	<0.001	Yes	Yes
1854	1883	-0.10	-0.52	<0.001	Yes	Yes
1824	1853	-0.35	-0.77	<0.001	Yes	Yes
1794	1823	-0.48	-0.90	<0.001	Yes	Yes
1764	1793	-0.24	-0.66	<0.001	Yes	Yes
1734	1763	-0.23	-0.65	<0.001	Yes	Yes
1704	1733	-0.32	-0.74	<0.001	Yes	Yes
1674	1703	-0.62	-1.04	<0.001	Yes	Yes
1644	1673	-0.28	-0.70	<0.001	Yes	Yes
1614	1643	-0.52	-0.94	<0.001	Yes	Yes
1584	1613	-0.79	-1.20	<0.001	Yes	Yes
1554	1583	-0.24	-0.66	<0.001	Yes	Yes
1524	1553	-0.19	-0.61	<0.001	Yes	Yes
1494	1523	-0.24	-0.66	<0.001	Yes	Yes
1464	1493	-0.29	-0.71	<0.001	Yes	Yes
1434	1463	-0.43	-0.85	<0.001	Yes	Yes
1404	1433	0.00	-0.42	0.002	Yes	No
1374	1403	-0.19	-0.61	<0.001	Yes	Yes
1344	1373	-0.38	-0.80	<0.001	Yes	Yes
1314	1343	-0.32	-0.74	<0.001	Yes	Yes
1284	1313	-0.34	-0.76	<0.001	Yes	Yes
1254	1283	-0.35	-0.77	<0.001	Yes	Yes
1224	1253	-0.05	-0.47	0.001	Yes	Yes
1194	1223	0.45	0.03	0.812	No	No
1164	1193	0.14	-0.28	0.050	No	No
1134	1163	0.30	-0.12	0.399	No	No
1104	1133	-0.54	-0.96	<0.001	Yes	Yes
1074	1103	-0.08	-0.50	0.001	Yes	No
1044	1073	-0.39	-0.81	<0.001	Yes	Yes
1014	1043	-0.16	-0.58	<0.001	Yes	Yes
984	1013	0.11	-0.31	0.029	Yes	No
954	983	0.36	-0.06	0.648	No	No
924	953	0.27	-0.15	0.316	No	No
894	923	0.04	-0.38	0.009	Yes	No
864	893	0.19	-0.23	0.123	No	No
834	863	-0.05	-0.47	0.001	Yes	No
804	833	-0.15	-0.57	<0.001	Yes	Yes
774	803	-0.14	-0.56	<0.001	Yes	Yes

744	773	0.10	-0.32	0.021	Yes	No
714	743	0.15	-0.27	0.040	No	No
684	713	-0.18	-0.60	0.000	Yes	Yes
654	683	-0.06	-0.48	0.001	Yes	Yes
624	653	-0.29	-0.71	<0.001	Yes	Yes
594	623	-0.35	-0.77	<0.001	Yes	Yes
564	593	-0.37	-0.79	<0.001	Yes	Yes
534	563	-0.64	-1.06	<0.001	Yes	Yes
504	533	-0.01	-0.43	0.002	Yes	No
474	503	0.08	-0.34	0.024	Yes	No
444	473	-0.22	-0.64	<0.001	Yes	Yes
414	443	-0.24	-0.66	<0.001	Yes	Yes
384	413	-0.17	-0.59	<0.001	Yes	Yes
354	383	-0.19	-0.61	<0.001	Yes	Yes
324	353	-0.45	-0.87	<0.001	Yes	Yes
294	323	-0.31	-0.73	<0.001	Yes	Yes
264	293	-0.10	-0.52	<0.001	Yes	Yes
234	263	0.13	-0.29	0.041	No	No
204	233	-0.06	-0.48	0.001	Yes	No
174	203	-0.03	-0.45	0.001	Yes	No
144	173	-0.25	-0.67	<0.001	Yes	< Yes
114	143	0.23	-0.19	0.143	No	No
84	113	0.13	-0.29	0.035	Yes	No
54	83	0.24	-0.18	0.185	No	No
24	53	0.56	0.14	0.336	No	No
-7	23	0.12	-0.30	0.046	No	No
-37	-8	0.03	-0.39	0.007	Yes	No
-67	-38	-0.04	-0.45	0.004	Yes	No
-97	-68	0.16	-0.26	0.071	No	No
-127	-98	-0.11	-0.53	<0.001	Yes	Yes

Table S11. As Table S5 but for the mean temperature anomaly of contiguous 30-year periods and the difference (Diff) from the last 30 years of the instrumental record using the CPS reconstruction.

CPS data: instrumental data for last 30 years						
start	End	Mean	Diff	t test	FDR	Bonf.
1986	2015	0.90				
1944	1973	0.19	-0.71	<0.001	Yes	Yes
1914	1943	0.07	-0.83	<0.001	Yes	Yes
1884	1913	-0.34	-1.24	<0.001	Yes	Yes
1854	1883	-0.13	-1.03	<0.001	Yes	Yes
1824	1853	-0.33	-1.23	<0.001	Yes	Yes
1794	1823	-0.46	-1.36	<0.001	Yes	Yes
1764	1793	-0.22	-1.12	<0.001	Yes	Yes
1734	1763	-0.17	-1.07	<0.001	Yes	Yes
1704	1733	-0.36	-1.26	<0.001	Yes	Yes
1674	1703	-0.59	-1.49	<0.001	Yes	Yes
1644	1673	-0.31	-1.21	<0.001	Yes	Yes
1614	1643	-0.49	-1.39	<0.001	Yes	Yes
1584	1613	-0.79	-1.69	<0.001	Yes	Yes
1554	1583	-0.24	-1.14	<0.001	Yes	Yes
1524	1553	-0.19	-1.09	<0.001	Yes	Yes
1494	1523	-0.18	-1.08	<0.001	Yes	Yes
1464	1493	-0.40	-1.30	<0.001	Yes	Yes
1434	1463	-0.41	-1.31	<0.001	Yes	Yes
1404	1433	0.01	-0.89	<0.001	Yes	Yes
1374	1403	-0.29	-1.19	<0.001	Yes	Yes
1344	1373	-0.47	-1.37	<0.001	Yes	Yes
1314	1343	-0.36	-1.26	<0.001	Yes	Yes
1284	1313	-0.47	-1.37	<0.001	Yes	Yes
1254	1283	-0.50	-1.40	<0.001	Yes	Yes
1224	1253	-0.34	-1.24	<0.001	Yes	Yes
1194	1223	-0.03	-0.93	<0.001	Yes	Yes
1164	1193	-0.03	-0.93	<0.001	Yes	Yes
1134	1163	-0.08	-0.98	<0.001	Yes	Yes
1104	1133	-0.64	-1.54	<0.001	Yes	Yes
1074	1103	-0.24	-1.14	<0.001	Yes	Yes
1044	1073	-0.26	-1.16	<0.001	Yes	Yes
1014	1043	-0.29	-1.19	<0.001	Yes	Yes
984	1013	-0.03	-0.93	<0.001	Yes	Yes
954	983	0.11	-0.79	<0.001	Yes	Yes
924	953	0.18	-0.72	<0.001	Yes	Yes
894	923	-0.08	-0.98	<0.001	Yes	Yes
864	893	0.14	-0.76	<0.001	Yes	Yes
834	863	-0.13	-1.03	<0.001	Yes	Yes

804	833	-0.28	-1.18	<0.001	Yes	Yes
774	803	-0.21	-1.11	<0.001	Yes	Yes
744	773	0.26	-0.64	<0.001	Yes	Yes
714	743	0.19	-0.71	<0.001	Yes	Yes
684	713	-0.11	-1.02	<0.001	Yes	Yes
654	683	-0.06	-0.96	<0.001	Yes	Yes
624	653	-0.25	-1.15	<0.001	Yes	Yes
594	623	-0.26	-1.16	<0.001	Yes	Yes
564	593	-0.29	-1.19	<0.001	Yes	Yes
534	563	-0.55	-1.45	<0.001	Yes	Yes
504	533	0.00	-0.90	<0.001	Yes	Yes
474	503	0.06	-0.84	<0.001	Yes	Yes
444	473	-0.20	-1.10	<0.001	Yes	Yes
414	443	-0.27	-1.17	<0.001	Yes	Yes
384	413	-0.22	-1.12	<0.001	Yes	Yes
354	383	-0.25	-1.15	<0.001	Yes	Yes
324	353	-0.46	-1.36	<0.001	Yes	Yes
294	323	-0.28	-1.18	<0.001	Yes	Yes
264	293	-0.06	-0.96	<0.001	Yes	Yes
234	263	0.23	-0.67	<0.001	Yes	Yes
204	233	-0.05	-0.95	<0.001	Yes	Yes
174	203	0.01	-0.89	<0.001	Yes	Yes
144	173	-0.26	-1.16	<0.001	Yes	Yes
114	143	0.26	-0.64	<0.001	Yes	Yes
84	113	0.15	-0.75	<0.001	Yes	Yes
54	83	0.29	-0.61	<0.001	Yes	Yes
24	53	0.55	-0.35	0.015	Yes	No
-7	23	0.08	-0.82	<0.001	Yes	Yes
-37	-8	-0.05	-0.95	<0.001	Yes	Yes
-67	-38	-0.08	-0.98	<0.001	Yes	Yes
-97	-68	0.29	-0.61	<0.001	Yes	Yes
-127	-98	-0.10	-1.00	<0.001	Yes	Yes

Table S12. As Table 11 but for the BHM reconstruction.

BHM reconstruction: instrumental data for last 30 years						
start	End	Mean	Diff	t test	FDR	Bonf.
1986	2015	0.90				
1944	1973	0.27	-0.63	<0.001	Yes	Yes
1914	1943	0.20	-0.70	<0.001	Yes	Yes
1884	1913	-0.34	-1.24	<0.001	Yes	Yes
1854	1883	-0.10	-1.00	<0.001	Yes	Yes
1824	1853	-0.35	-1.25	<0.001	Yes	Yes
1794	1823	-0.48	-1.38	<0.001	Yes	Yes
1764	1793	-0.24	-1.14	<0.001	Yes	Yes
1734	1763	-0.23	-1.13	<0.001	Yes	Yes
1704	1733	-0.32	-1.22	<0.001	Yes	Yes
1674	1703	-0.62	-1.52	<0.001	Yes	Yes
1644	1673	-0.28	-1.18	<0.001	Yes	Yes
1614	1643	-0.52	-1.42	<0.001	Yes	Yes
1584	1613	-0.79	-1.69	<0.001	Yes	Yes
1554	1583	-0.24	-1.14	<0.001	Yes	Yes
1524	1553	-0.19	-1.09	<0.001	Yes	Yes
1494	1523	-0.24	-1.14	<0.001	Yes	Yes
1464	1493	-0.29	-1.19	<0.001	Yes	Yes
1434	1463	-0.43	-1.33	<0.001	Yes	Yes
1404	1433	0.00	-0.90	<0.001	Yes	Yes
1374	1403	-0.19	-1.09	<0.001	Yes	Yes
1344	1373	-0.38	-1.28	<0.001	Yes	Yes
1314	1343	-0.32	-1.22	<0.001	Yes	Yes
1284	1313	-0.34	-1.24	<0.001	Yes	Yes
1254	1283	-0.35	-1.25	<0.001	Yes	Yes
1224	1253	-0.05	-0.95	<0.001	Yes	Yes
1194	1223	0.45	-0.45	<0.001	Yes	Yes
1164	1193	0.14	-0.76	<0.001	Yes	Yes
1134	1163	0.30	-0.60	<0.001	Yes	Yes
1104	1133	-0.54	-1.44	<0.001	Yes	Yes
1074	1103	-0.08	-0.98	<0.001	Yes	Yes
1044	1073	-0.39	-1.29	<0.001	Yes	Yes
1014	1043	-0.16	-1.06	<0.001	Yes	Yes
984	1013	0.11	-0.79	<0.001	Yes	Yes
954	983	0.36	-0.54	<0.001	Yes	Yes
924	953	0.27	-0.63	<0.001	Yes	Yes
894	923	0.04	-0.86	<0.001	Yes	Yes
864	893	0.19	-0.71	<0.001	Yes	Yes
834	863	-0.05	-0.95	<0.001	Yes	Yes
804	833	-0.15	-1.05	<0.001	Yes	Yes
774	803	-0.14	-1.04	<0.001	Yes	Yes

744	773	0.10	-0.80	<0.001	Yes	Yes
714	743	0.15	-0.75	<0.001	Yes	Yes
684	713	-0.18	-1.08	<0.001	Yes	Yes
654	683	-0.06	-0.96	<0.001	Yes	Yes
624	653	-0.29	-1.19	<0.001	Yes	Yes
594	623	-0.35	-1.25	<0.001	Yes	Yes
564	593	-0.37	-1.27	<0.001	Yes	Yes
534	563	-0.64	-1.54	<0.001	Yes	Yes
504	533	-0.01	-0.91	<0.001	Yes	Yes
474	503	0.08	-0.82	<0.001	Yes	Yes
444	473	-0.22	-1.12	<0.001	Yes	Yes
414	443	-0.24	-1.14	<0.001	Yes	Yes
384	413	-0.17	-1.07	<0.001	Yes	Yes
354	383	-0.19	-1.09	<0.001	Yes	Yes
324	353	-0.45	-1.35	<0.001	Yes	Yes
294	323	-0.31	-1.21	<0.001	Yes	Yes
264	293	-0.10	-1.00	<0.001	Yes	Yes
234	263	0.13	-0.77	<0.001	Yes	Yes
204	233	-0.06	-0.96	<0.001	Yes	Yes
174	203	-0.03	-0.93	<0.001	Yes	Yes
144	173	-0.25	-1.15	<0.001	Yes	Yes
114	143	0.23	-0.67	<0.001	Yes	Yes
84	113	0.13	-0.77	<0.001	Yes	Yes
54	83	0.24	-0.66	<0.001	Yes	Yes
24	53	0.56	-0.34	0.009	Yes	No
-7	23	0.12	-0.78	<0.001	Yes	Yes
-37	-8	0.03	-0.87	<0.001	Yes	Yes
-67	-38	-0.04	-0.94	<0.001	Yes	Yes
-97	-68	0.16	-0.74	<0.001	Yes	Yes
-127	-98	-0.11	-1.01	<0.001	Yes	Yes

Comparison between the new European summer temperature reconstruction and independent instrumental and low and high-resolution proxy based summer temperature estimates from Europe

We compare the mean and the gridded BHM-based reconstruction with long independent station temperature series (Table S13; Figs. S7, S8) and summer temperature reconstructions from various proxy records (Table S13, Figs. S7, S9) which were not used for our reconstruction. In this context we evaluate the reconstruction skill by taking the results of the predictive run of the BHM reconstruction, estimated at the locations of the independent temperature data. Using all ensemble members, we calculate the Pearson's correlation coefficient and the standard deviation ratio (SDR) between the ensemble members of the predictive experiment and the independent summer temperature data prior to 1850 CE, the beginning of the CRUTEM 4v instrumental data (Jones et al. 2012). These are eventually averaged over all ensemble members. Where the temporal resolution of the reconstruction did not match that of the considered independent proxy data, the reconstruction was resampled and time averaged to match both the resolution and the averaging procedures used in these studies. Results are provided in Table S5 and shown in Figs. S8 and S9. Correlation coefficients are high between our reconstructed summer temperatures and long instrumental data from the Alps and Sweden close to proxy locations (Table S13; Fig. S8). Correlation coefficients with instrumental data from the Netherlands, Central England and Northern Italy are smaller but still statistically significant ($p < 0.05$ level; Table S13; Fig. S8). In the earliest periods a number of series were composed of readings of imperfect instruments located in various locations and exposures. For instance the earliest part of the CET series till 1672 is less reliable being composed of multi-proxy in combination with extrapolation from various readings (Parker et al, 1992). The subsequent instrumental period of CET is composed of indoor observations, as recommended by the protocol of the Royal Medical Society, London (Jurin 1723) for three main reasons: to use non-weatherproof instruments, to smooth out the diurnal cycle, and for health purposes. In Europe, this recommendation was followed by the observers that adhered to the Royal Network, flourished 1724-1735. In the second half of the 18th century, the interest for health purposes was enhanced with the Network of the Société Royale de Médecine, Paris that flourished in 1776-1786. Outdoor observations with weather and climate purposes returned with the Societas Meteorologica Palatina, Mannheim, flourished 1781-1792. The indoor-outdoor change was responsible for dishomogeneity in the last quarter of the 18th century (Camuffo, 2002). Another common bias prior to about 1850 CE is for instrumental summer temperatures generally higher than the reconstructions, though they are still within the uncertainty range (Fig. S8 b,c,g). This can be attributed to the well-known summer temperature warm bias due to missing Stevenson shielding (Moberg et al. 2003; Frank et al. 2007; Böhm et al. 2010). However, not all series were affected by this bias. In Italy the tradition was with a thermometer on the shade, hung on a North facing wall, following a recommendation of the Medici Network, Florence, active from 1655 to 1670 (Camuffo and Bertolin, 2012a). It has been suggested (Middleton, 1966) that the solar screen was applied in 1830, but in Padua since 1780 Toaldo shielded his thermometer against the direct radiation of the summer sunrise and sunset, although the thermometer was hung outside of a North-facing window (Camuffo, 2002). In conjunction with the standard deviation ratio (SDR), the well-captured temperature amplitude between the early 19th century cold episode and the late 20th century warming, and the good qualitative agreement with the Tornedalen and Histalp temperature series (Table S13) indicate that inter-annual to decadal variability is very well preserved at least in areas that are close to proxy locations, while more data are needed in other regions where the skill is lower (e.g. British Isles).

Table S13: Compilation of independent summer temperature reconstructions and long instrumental series used for comparisons with the BHM reconstruction. Mean correlation coefficients (* means significant at $p < 0.05$ level) and mean standard deviation ratios (SDR) between reconstructed summer temperatures and independent proxy/instrumental data are shown together with coordinates and data type. ** not indicated in map, see Figs. S10 and S11

No., Region	East	North	r	SDR	Type	Period	Reference
1, CET	-1.5	52.7	0.28*	0.974	Long instrumental data	1659-2007	Manley (1974, updated)
2, N-Italy	11	45	0.35*	1.032	Long instrumental data	1655-2007 (with gaps)	Camuffo and Bertolin (2012a,b); Camuffo et al. (2010)
3, De Bilt	5.18	52.11	0.39*	0.887	Long instrumental data	1706-2003	van Engelen and Nellestin 1996, updated
4, Stockholm	18.0	59.3	0.42*	0.873	Long instrumental data	1756-2010	Moberg and Bergström, 1997; Moberg et al. 2002
5, Histalp	11.7	46.3	0.77*	0.907	Long instrumental data	1774-2007	Böhm et al. (2009)
6, Tornedalen	24	66	0.71*	0.993	Long instrumental data	1802-2008	Klingbjer and Moberg (2003)
7, SE England	1	52.6	0.03	1.220	Grain harvest date (AMJJ temperature)	1256-1431	Pribyl et al. (2012)
8, Vestlandet	7	61	0.49*	1.163	Multi-proxy reconstruction (spring to summer)	1732-2003	Nordli et al. (2003)
9, Finland	25.4	60.3	0.24	0.606	Chironomid temperatures, lake sediment, irregular multidecadal resolution	1330-2000	Luoto et al. (2009)
10, Swiss Alps	9.76	46.43	0.16	0.704	Chironomid July temperature (3y average)	1780-1998	Larocque-Tobler et al. (2010)
10, Swiss Alps	9.76	46.43	0.4*	0.416	Chironomid July temperature (20y average)	1120-1990	Larocque-Tobler et al. (2010)
11, Alps	4-19	43-49	0.6*	1.034	Multi-proxy summer reconstruction (not fully independent, as it includes Büntgen et al. 2006) data	1067-1996	Trachsel et al. (2012)
12, Russian Plains	30-40	53-58	0.33*	1.693	Multi-proxy summer T reconstruction	605-1995	Sleptsov and Klimenko (2003)
13, Finland	30.1 2	64.28	0.15	1.129	Temperature, lake sediment irregular multi-decadal resolution	470-1990	Luoto and Helama (2010)
14, Northern Spain	-4	43	0.117	3.022	Speleothem (low resolution)	-138-2003	Martín-Chivelet et al. (2011)
**15, Western Europe	-10-40	35-70	0.1	0.910	Multi-proxy spring/summer T reconstruction	600-2007	Guiot et al. (2010)
**16, Europe	-25-40	35-70	0.41*	1.27	Multi-proxy summer T reconstruction	1500-1899	Luterbacher et al. (2004)

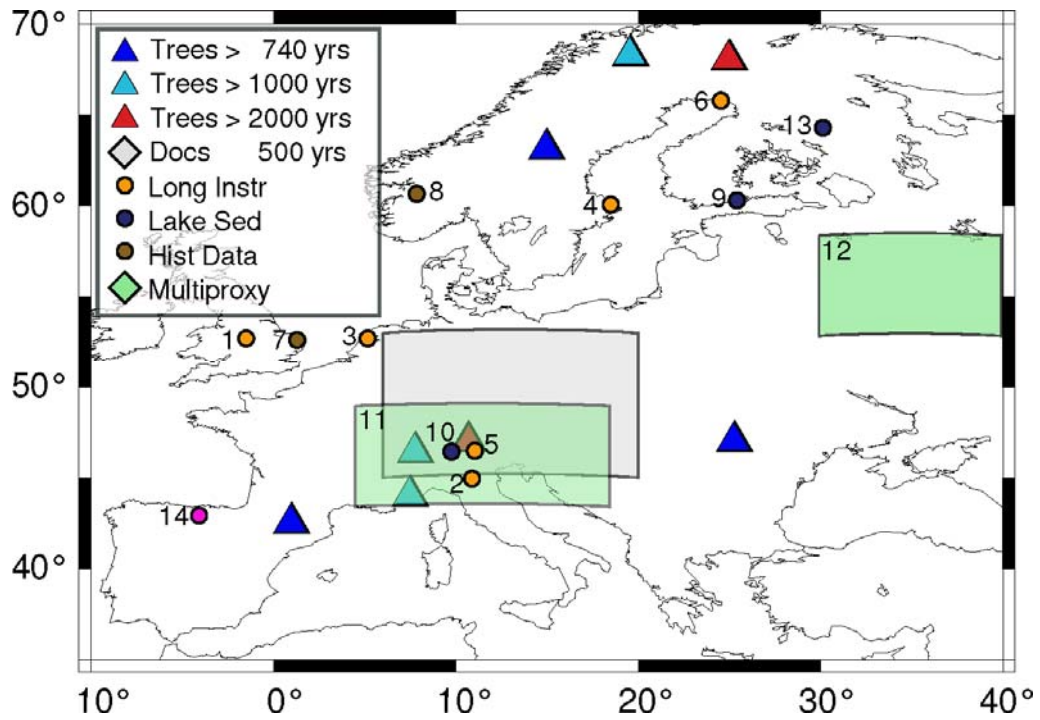


Figure S7: Locations of proxy data in this study (triangles for tree ring data and grey shaded area for documentary evidence, see Fig. 1) and independent local (circles) and regional (green shaded area) summer temperature evidence. See Table S13 for details.

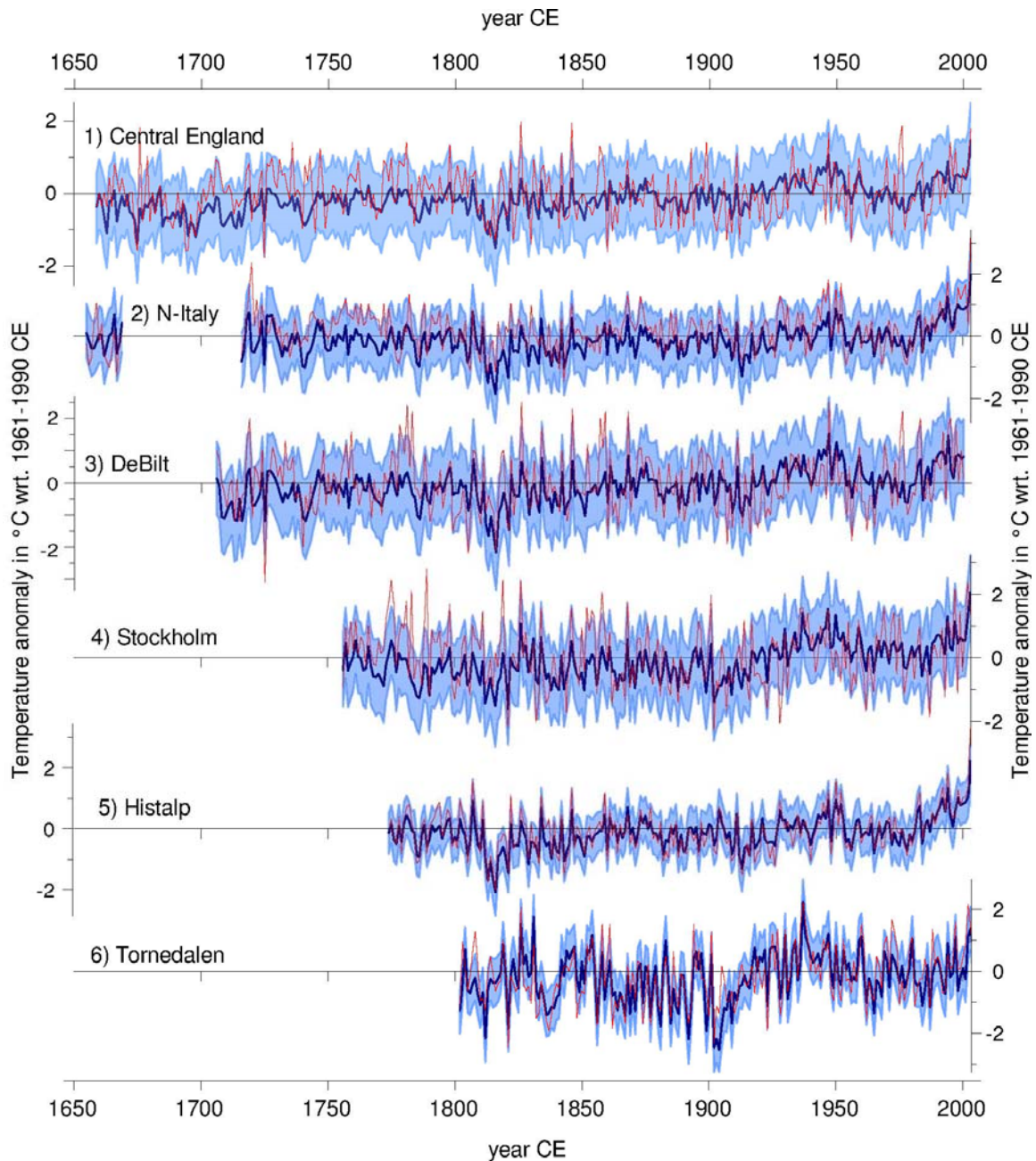


Figure S8: Comparison of reconstructed BHM-based summer temperature anomalies (blue line) and uncertainties (shading) with independent summer temperature records (red lines) from different areas of Europe. All anomalies are computed with respect to 1961-1990 CE (see Table S13 and Fig. S7 for details).

The temperatures reconstructed from grain harvest data from the southeastern UK (Pribyl et al. 2012) bear little resemblance to our reconstruction (Fig. S9) but lie within the uncertainty range. Lake sediment data from Finland (Luoto et al., 2009; Luoto and Helama 2010) behave similarly to the reconstruction at decadal to multi-decadal time-scales, but minima and maxima of the longer time-series (Luoto and Helama 2010) are shifted by several decades, possibly indicating problematic age-depth models in these particular sediment data (Fig. S9). The reconstructed temperatures at Lake Silvaplana in the Alps were filtered with triangular

filters (three and twenty years) to be more readily comparable with the chironomid-inferred temperature reconstruction from the lake sediments. While agreement of the data over the 20th century is good, coherence is low for the early 19th century cooling episode (Fig. S9). The 20-yr filtered lake sediment data indicates relative warming during the 11th century that is absent in our reconstructions, at least when using the same type of filter. The BHM reconstruction compares quite well with the decadal summer temperature reconstruction from the Russian Plain except for the early 19th century cooling, even though the BHM reconstruction does not include any proxy data from the Russian Plains target area 30-40° E 53-58° N (Fig. S9). The match with the low-resolution speleothem record from northern Spain is also quite good during the first millennium period 300-1000 CE (it should be noted that this record is the only reconstruction available for comparison during this period). The comparisons with these independent decadal-resolved proxy-based reconstructions indicate that the BHM reconstruction faithfully preserved decadal-to-centennial variability (Fig. S9).

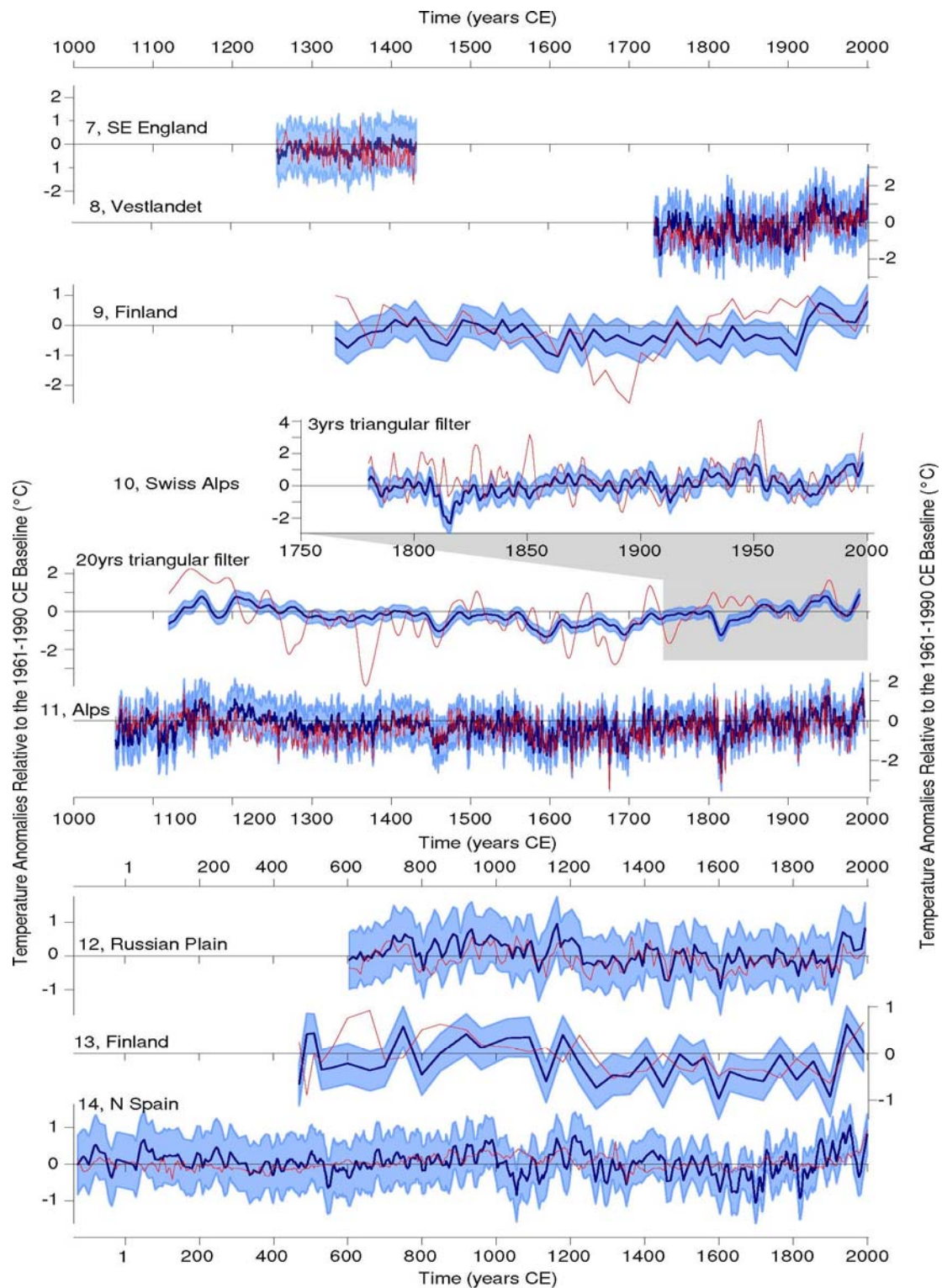


Figure S9: Comparison of reconstructed BHM-based summer temperature anomalies (blue line) and uncertainties (shading) with long independent summer temperature reconstructions (red lines) from proxies over different areas of Europe, all anomalies with respect to 1961-1990 CE, see Table S13 and Fig. S7 for details.

We find good agreement with the independent European mean and gridded summer temperature reconstructions by Luterbacher et al. (2004) covering the past five centuries

(Table S13, Fig. S10). However, there is strong indication that the spatial mean of the earlier reconstruction has less low-frequency variability than the BHM reconstruction. The correlations over most of the grid are statistically significant, except for the southeastern part (Fig. S10), where no proxy data are available and the uncertainties are largest for both reconstructions. The ensemble mean of the SDR is displayed in Table S13. We also compare our reconstruction with the mean and gridded April-September temperature reconstruction by Guiot et al. (2010). The correlations between both data sets are generally very low (Table S13) though statistically significant in central Europe (Fig. S11). Some of the discrepancies are likely due to the different target seasons and match those found comparing summer (JJA) with warm season (May-October) averages in the instrumental CRUTEM data (Jones et al. 2012, not shown).

Many established reconstruction methods suffer from a loss of reconstructed temporal variability when reconstructing large-scale area means. For spatially resolved climate field reconstructions this effect is more pronounced (von Storch et al. 2004, Bürger et al. 2006; Christiansen et al. 2009; Smerdon and Kaplan, 2007, Smerdon et al. 2011; Christiansen 2011), although it is not present in spatially-resolved annual and February-March temperature reconstructions in western temperate North America (Wahl and Smerdon (2012) and a spatially-resolved precipitation reconstruction for California and western Nevada in the United States (Diaz and Wahl, 2015). However, BHM as implemented here has been shown to skilfully reconstruct much of the target variance in pseudo-proxy tests over Europe (Werner et al. 2012) and North America (Tingley and Huybers 2010a,b) and to estimate confidence intervals that faithfully include the target data. The above presented comparisons to data suggest that the results from pseudoproxy experiments are also applicable for real world applications.

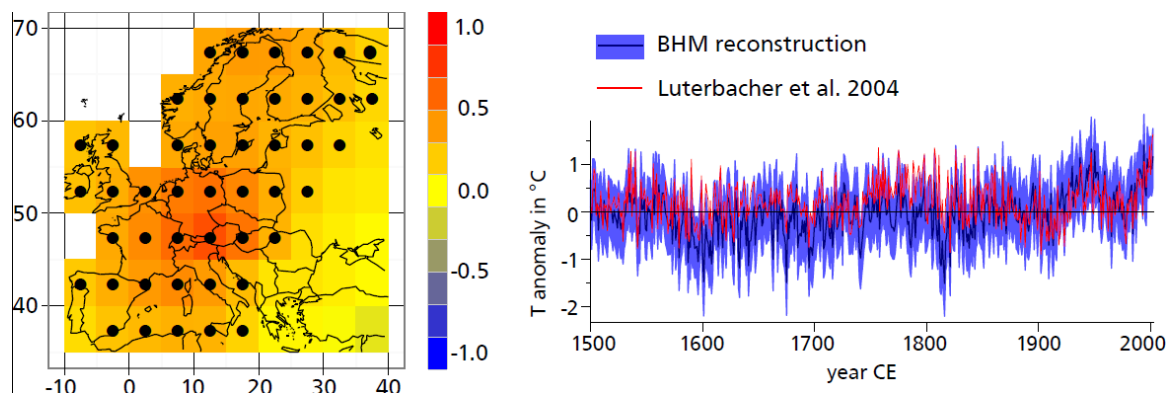


Figure S10: Correlation at each grid cell between Luterbacher et al. (2004) gridded -summer reconstructions and the BHM JJA estimates. The black dots mark the locations where correlations are significant at $p < 0.05$ level. Right: Comparison of area weighted averaged European temperature anomalies, red: Luterbacher et al. (2004), blue: mean BHM reconstruction.

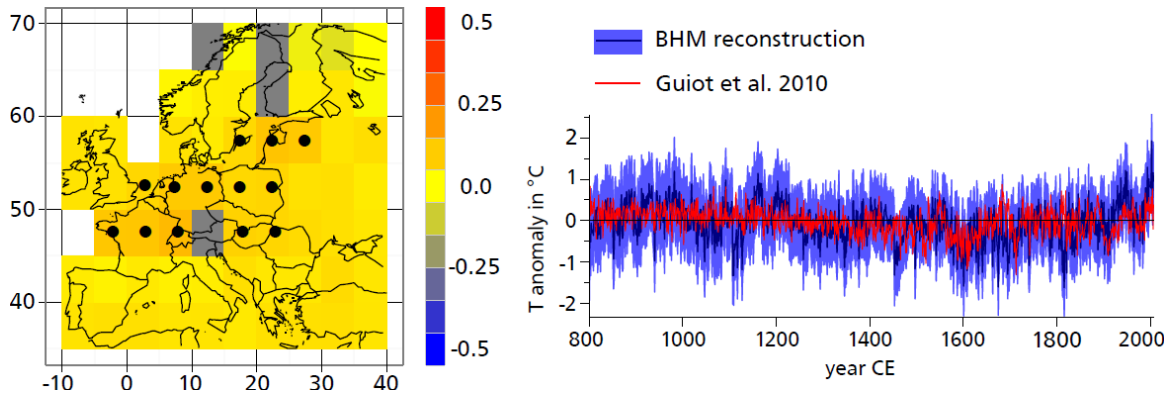


Figure S11: Left: Correlation at each grid cell between Guiot et al. (2010) gridded April-September reconstructions and the BHM JJA estimates. The black dots mark the locations where correlations are significant at $p < 0.05$ level. The Guiot et al. (2010) reconstruction is not available within the grey shaded grids. Right: Comparison of area-weighted averaged European temperature anomalies, red: Guiot et al. (2010), blue: mean BHM reconstruction.

Details on the computation of Figure 2, main text

For each grid point and ensemble member, the BHM reconstructed summer temperature of each year is expressed as anomaly with respect to the 755-2003 CE period. 11-yr running centered means of the anomalies are then calculated at each grid point for 11-yr periods with complete data (i.e. from 755-765 CE to 1993-2003 CE). For each ensemble member, the local maximum and minimum 11-yr mean anomaly of 755-2003 CE is chosen and the corresponding ensemble mean is represented in Fig. 2 by the height of the bars and filled shading with contour interval of $+0.2$ °C starting at 0 °C. The height of each flag above the bar represents the associated temperature uncertainty at the grid point scale, defined as $+2$ standard deviations (SD) of the ensemble distribution of the warmest 11-yr mean temperatures.

To estimate the most likely date of the warmest 11-yr mean period across the ensemble, we constructed the frequency distribution of dates, taking the central year of the warmest 11-yr mean period of each member. By using centred running temporal windows of 100-yr width, we calculated the number of members having their warmest and coolest 11-yr mean period in the same 100-yr interval. The most populated bin is the most likely date of the warmest and coolest 11-yr mean period across the ensemble, and is indicated in the top of each grid point by coloured squared symbols, according to the colour bar. Flags marked with a black square denote grid points where more than 75% of the ensemble members have the warmest (or coolest) 11-yr mean period in the same 100-yr bin (i.e. agree on the timing of the warmest and coolest period). Finally, for each ensemble member, the time series of the 11-yr mean summer temperature anomalies at each grid point, T_i , were normalized by dividing by the standard deviation $SD = \sqrt{\sum (T_i - T)^2 / N - 1}$ where N equals the number of all available 11-yr periods of the 755-2003 CE record, and T is the local 755-2003 CE average of the 11-yr mean temperature series T_i . For each 11-yr period, we searched for those grid points with temperature anomalies above and below the local 2 SD value and calculated the percentage of the European area that they represent. This allows the construction of the time-evolving percentage of European area with 11-yr mean temperatures above and below 2 SD for each ensemble member. The light (dark) red shading in the front plot of the map represents the ensemble spread, defined as the 5th-95th percentile range (± 0.5 sigma interval). The bottom panel of Fig. 2 (main text) is the same as the top panel but for 51-yr mean periods.

We also explored the degree to which the 20th century summer temperatures in Europe was exceptional (Fig. S12). To do so, we selected for each ensemble member and each grid point the warmest 11-yr mean period of the 20th century (T_m) and compared it with all 11-yr periods of the pre-1900 interval. We calculated the last time that temperature was reached in the past, and the time interval between them. If T_m is already the warmest local temperature of the entire 755-2003 CE period, the corresponding duration is computed with respect to the beginning of the series. For each grid point, the ensemble distribution of durations was calculated for running centred 100-yr bins and the most populated bin of duration was selected. The corresponding durations are shown with colour shading in Fig. S12, and the size of the black dot is proportional to the degree of agreement across the ensemble members at the given grid point (i.e. the percentage of ensemble members in that bin).

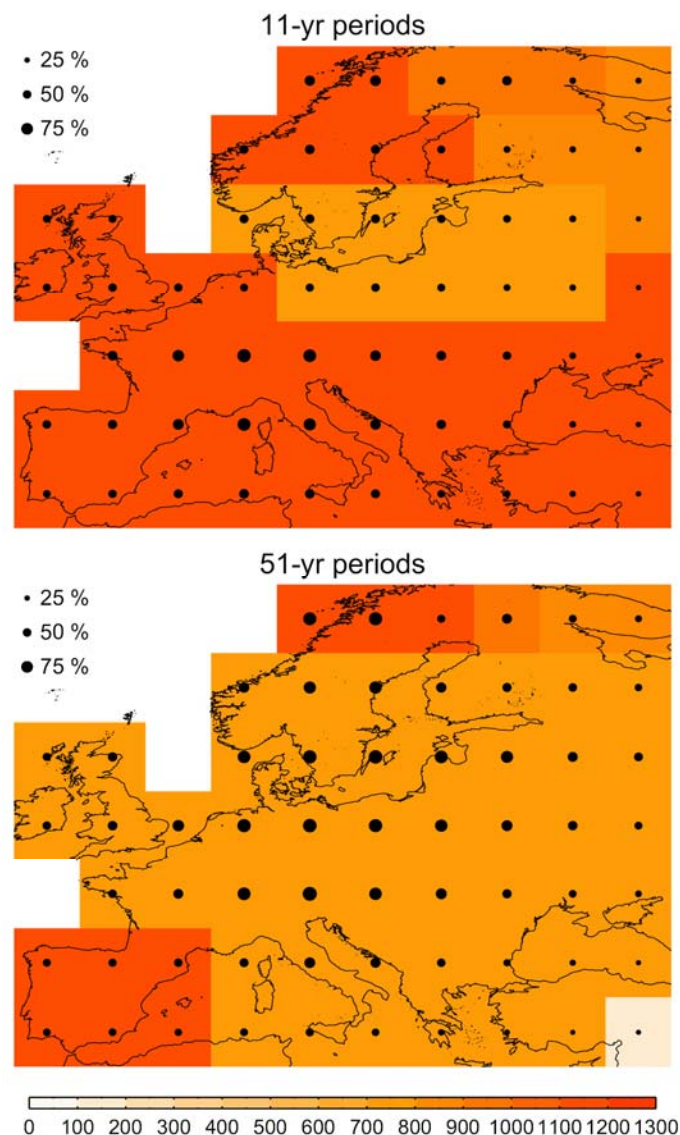


Figure S12: Top: spatial distribution with the time interval (in years) between the last 11-yr mean period of the pre-1900 CE period with equal or higher temperatures than the warmest 11-yr mean period of the 20th century. The black dots denote the agreement across the ensemble, with the size proportional to percentage of members giving the same levels of precedence. Bottom: As in the top panel but for 51-yr mean periods.

GCM simulations over the last 2000 years and model data comparison

Atmosphere-ocean coupled general circulation model (AOGCM) simulations are compared to the European summer temperature reconstructions at the continental and grid point scale (Figs. 3 and 4 in the main text). AOGCMs show their highest skill at larger scales and present reduced reliability at the grid-point scale, as a consequence of factors such as simplified local orography and surface physics (e.g. von Storch 2004). Therefore, we discuss the grid point scale comparison (Fig. 4 in the main text) of AOGCM output and the climate reconstructions only on regional scales. Near surface temperatures at these spatial scales indicate a large spatial coherence (Jones et al. 1997; Büntgen et al. 2010; Ljungqvist et al. 2012) due to the influence of large-scale circulation and external forcings (Hegerl et al. 2011). A growing number of climate simulations from comprehensive state-of-the-art AOGCMs and Earth System Models (ESMs) extend well back into the first millennium CE. These models are driven with different estimates of external climate forcings (Fernández-Donado et al. 2013). We consider an ensemble of paleoclimate simulations including 37 experiments. 26 of these simulations were performed with eight different AOGCMs under diverse external forcing configurations (Table S14; Fernández-Donado et al. 2013). The remaining eleven simulations are part of the Paleo Model Intercomparison Project Phase III (PMIP3; Taylor 2009; Braconnot et al. 2012) and generally follow the PMIP3-protocol for estimates of external forcing factors (Schmidt et al. 2011, 2012). PMIP3 simulations are made available from the Earth System Grid (<http://cmip-pcmdi.llnl.gov/index.html>). The model ensemble used herein matches that used in Masson-Delmotte et al. (2013).

The model experiments used for comparison differ in model complexity, in the forcing reconstructions used and their implementation (see Table S14). Model resolution in the atmosphere ranges from R21 spectral resolution (Phipps et al. 2011) to a maximum grid resolution of $1.25^{\circ}\text{lat} \times 0.9^{\circ}\text{lon}$, similar to those used in the CMIP5 historical and future scenario simulations (e.g. Landrum et al. 2013). The most notable differences in the forcing configurations (Schmidt et al. 2011; Fernández-Donado et al. 2013) concern long-term variations in total solar irradiance (TSI). Most of the earlier TSI-reconstructions exhibit a TSI-change larger than 0.23 % from the Late Maunder Minimum (LMM) to present. The reassessment of the reconstructions by Lean et al. (2002) led to generally lower (about 0.05-0.1%) estimates (see Fernández-Donado et al. 2013). The PMIP3 simulations therefore adopt TSI-changes from the LMM to present on the order of 0.1 % (Schmidt et al. 2011, see also Schmidt et al. 2012). The coordinated convention for the forcing scenarios for the PMIP3 forcing ensemble allows flexible options for uncertain climate forcings for which multiple estimates are available. This results in differences in the setups between individual simulations. For example, the volcanic forcing data by Gao et al. (2008, 2012) incorporates larger changes than the data by Crowley and Unterman (2013). Some of the model-systems include an interactive carbon cycle but only the ECHAM5/MPIOM ensemble performs a prognostic calculation of atmospheric CO₂. The other simulations are driven by prescribed atmospheric concentrations from reconstructions. While the interactively calculated CO₂ concentrations in ECHAM5/MPIOM show smaller variations than reconstructions on decadal to centennial time scales (JungCLAUS et al., 2010), we consider this a minor distinction. Newer simulations apply land-use changes as additional forcing, which are missing in most of the pre-PMIP (except for EH5/MPIOM) simulations. Note that Ahn et al. (2011) published a new estimate of atmospheric CO₂ concentrations over the last millennium that is not included in the PMIP3-protocol (Schmidt et al., 2011, 2012).

Table S14: Models and experiments considered for the analysis (column 1); horizontal and vertical resolution of atmospheric and ocean model components (columns 2 and 3); set of external forcings considered in the experiment configuration (column 4); number of simulations and length (column 5) and original reference describing the experiments (column 6). Legend for external forcing configuration: (W) solar forcing using stronger changes in amplitude (i.e. larger than 0.23 TSI change since LMM to present); (N) solar forcing using weaker changes in amplitude (i.e. lower than 0.1 % TSI change since LMM to present); (V) volcanic activity; (G) greenhouse gases; (A) anthropogenic aerosols; (L) land use changes; and (O) orbital variations.

Model	Atmosphere	Ocean	Forcings	Simulations	Reference
	Resolution/vertical levels			(n° runs/length)	
CCSM3	T31/18	3.6X2.8/25	WVG	1000-2100 CE (4) 1500-2100	Hofer et al (2010)
CNRM	T42/31	2X2/31	WVGAL	(1) 1001-1999 CE	Swingedouw et al. (2010)
CSIRO	R21/18	2.8X1.6/21	NGO NVGO	(3)1-2000 CE (3)501-2000 CE	Phipps et al. (2013)
CSM1.4	T31/18	3.6X1.8/25	WVGA	(1) 850-1999 CE	Ammann et al. (2007)
ECHAM5-MPIOM	T31/19	3X3/40	NVGALO WVGALO	E1 (5) 800-2000 CE E2 (3) 800-2000 CE	Jungclaus et al. (2010)
ECHO-G	T30/19	2.8X2.8/20	WVG WGO	(2)1000-1990 CE (1) 8000 - 0 BP	González-Rouco et al. (2006) Wagner et al. (2007)
HadCM	3.75X2.5/19	1.25X1.25/20	WVGALO	(1) 1492-1999 CE	Tett et al. (2007)
IPSL	3.75x2.5/19	2x2/31	WGAO	(1) 1001-2000 CE	Servonnat et al. (2010)
PMIP3/CMIP5					
BCC	T42L26	tripolar, 1 lon x (1-1/3) lat, L40	Schmidt et al. (2011, 2012)	(1) 850-1850 CE	Wu (2012)
CCSM4	1.25x1.25L26	1.1 lon x 0.27-0.54 lat L60	Schmidt et al. (2011, 2012)	(1) 850-1850 CE	Landrum et al. (2013)
CSIRO	R21	2.8x1.6 L21	Schmidt et al. (2011, 2012)	(1) 851- 1850 CE	Phipps et al. (2013)
FGOALS	72x40L26	360x170L30	WVG	(1) 1000-1999 CE	Zhou et al. (2011)
GISS	2.5x2/L40	1x1.25xL32	Schmidt et al. (2011, 2012)	(3) 850-1850 CE	Schmidt et al. (2014) http://data.giss.nasa.gov/modelE/ar5/
HadCM	3.75X2.5/19	1.25x1.25L20	Schmidt et al. (2011, 2012)	(1) 800- 2000 CE	Schurer et al. (2013)
IPSL-CM5A-LR			Schmidt et al. (2011, 2012)	(1) 850-2005 CE	Khodri et al. (2015, in prep)
MIROC	T42L80	256x192 L44	Schmidt et al. (2011, 2012)	850- 1849 CE	Sueyoshi et al. (2013)
MPI-ESM	T63L47	GR15L40	Schmidt et al. (2011, 2012)	850-2005 CE	Jungclaus et al. (2014)

European summer temperature response after strong tropical volcanic eruptions

The European summer temperature response to the volcanic events is analysed for the PMIP3 model simulations through SEA (Fig. S13). Compared to the Crowley and Untermann (2012) series, the temperature response to the volcanic activity is stronger in the simulations driven by Gao et al. (2008) forcing (PAGES2k-PMIP3 group, 2015). Therefore, the temperature responses are separated according to the volcanic forcing presented in Table S14. Simulations driven by the Gao et al. (2008) forcing correspond to BCC and CCSM4, while CSIRO, HadCM3, MIROC, MPI and GISS ensemble use the Crowley and Untermann (2012) forcing (Table S14). For each volcanic forcing, the 12 strongest volcanic events are selected, following the same approach as in PAGES2k-PMIP3 group (2015). The temperature response of each simulation to each volcano is obtained as the spatial summer temperature difference between a given year and the average of the five years previous to the eruption (similar to the SEA of the reconstructions; Fig. S14). The simulated temperature response is analysed from the year of the eruption to the third year after the event. The standard deviation obtained from all the individual summer anomaly maps is shown as contours.

Figure S13 shows an overall European summer cooling as response to the volcanic events, larger for the year of the eruption and the year after. Simulations driven by the Gao et al. (2008) volcanic forcing show a stronger cooling compared to the Crowley and Untermann (2012) subgroup, in agreement with the results of PAGES2k-PMIP3 group (2015).

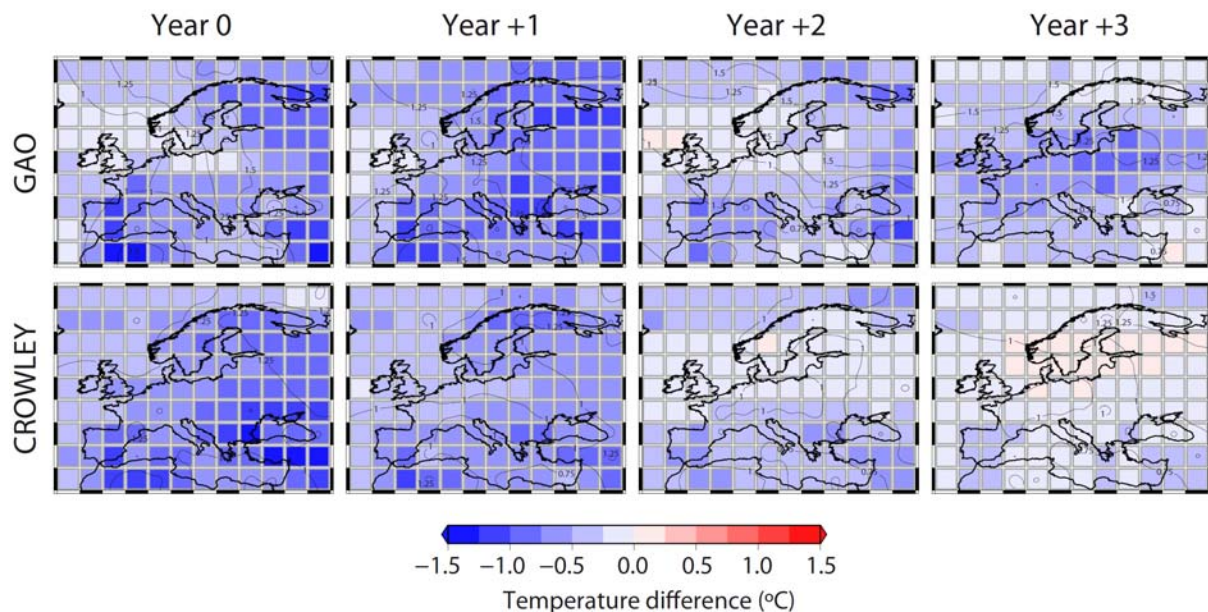


Figure S13. Superposed spatial composites of the summer temperature responses over the European region from PMIP3 simulations for their 12 strongest volcanic events over the past millennium. The temperature responses are shown from the year of the event to the third year after the eruption. Anomalies are presented with respect to the average of the five years previous to each eruption. Each map shows the average temperature difference (colours) from all the simulations driven by (top) Gao et al. (2008) and Crowley and Untermann (2012) (bottom) volcanic forcing. The standard deviations among all these temperature differences are also presented in the maps (contour lines).

Figure S14 shows the European summer temperature response to the volcanic events through SEA for the BHM reconstructions. The SEA is performed for the 13 strongest tropical volcanic eruptions (\geq VEI 5) published in Esper et al. (2013). The selected eruptions all occurred during the time period covered by the gridded BHM reconstruction.

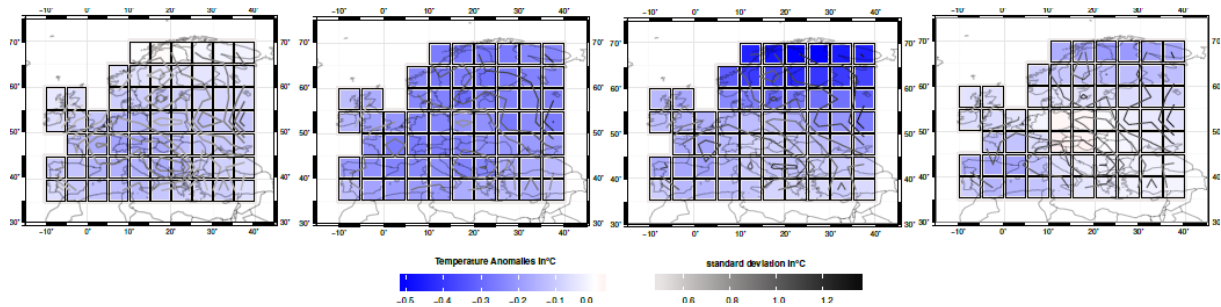


Figure S14. Superposed Epoch Analysis (SEA) of European summer temperature for the 13 strongest tropical volcanic eruptions (\geq VEI 5) published in Esper et al. (2013). The temperature responses ($^{\circ}$ C) are shown from the year of the event (left panel) to the third year after the eruption (right panel) with respect to the average of the five years preceding each eruption. Each map shows the average temperature difference (colours) from all ensemble members. The standard deviations among them are shown with contour lines

Figure S15 presents the European summer temperature response from the BHM reconstructions for a selection of strong tropical volcanic eruptions (Samalas 1257, Huaynaputina 1600, Parker 1641, Laki 1783/1784 and Tambora 1815). The responses (ensemble mean) are shown with respect to average temperatures over the five years preceding the eruption.

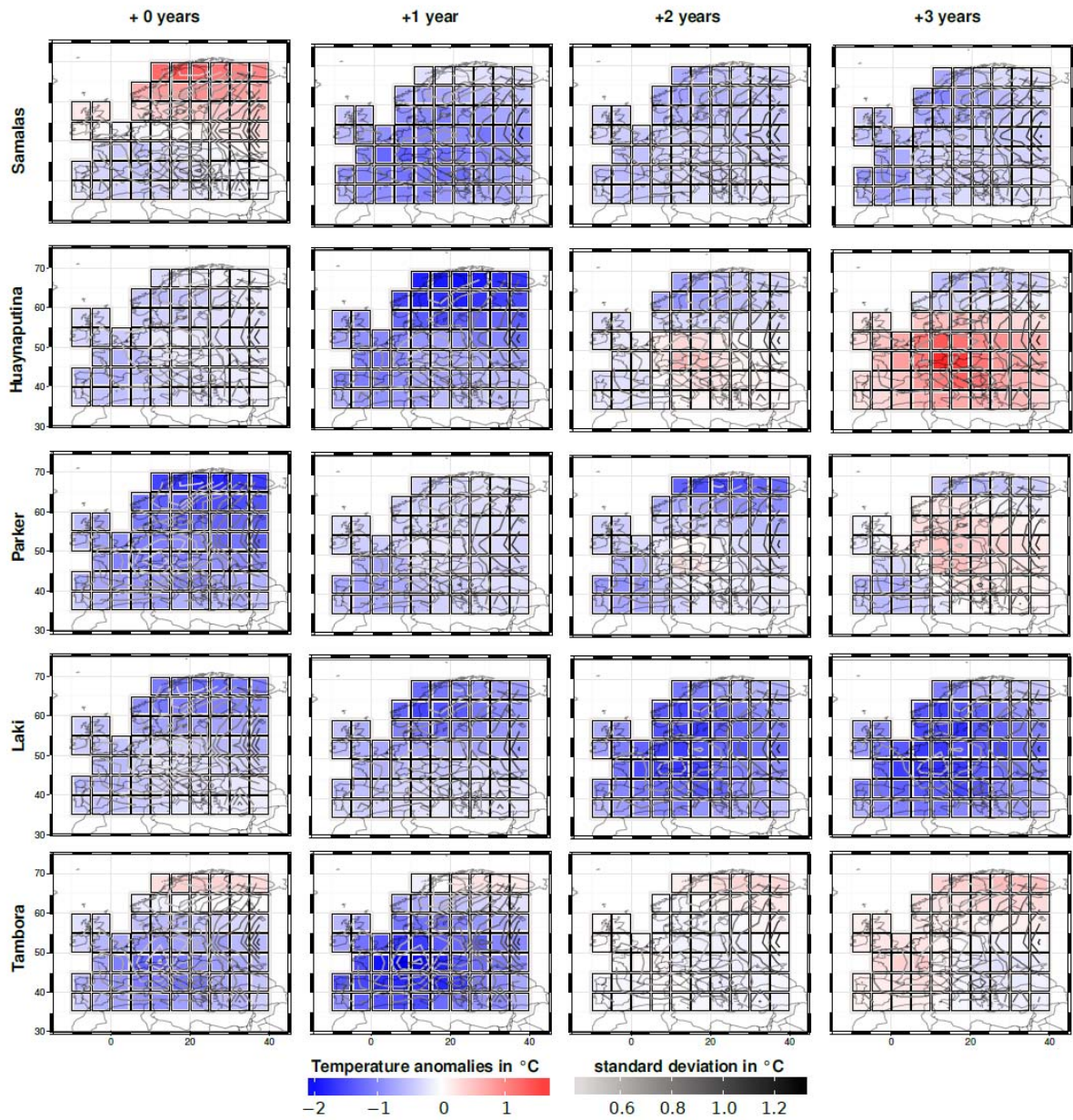


Figure S15. Summer temperature response ($^{\circ}\text{C}$) for a selection of strong volcanic eruptions (Samalas 1257, Huaynaputina 1600, Parker 1641, Laki 1783/1784 and Tambora 1815). The responses (ensemble mean) are shown with respect to average temperatures over the five years preceding the eruption. The standard deviations are calculated from the ensemble members and show an estimate of the uncertainty of the reconstructed response. Columns are the year of the eruption and the three following years

References SOM

- Ahn J, Brook E J, Mitchell L, Rosen J, McConnell J R, Taylor K, Etheridge D and Rubino M 2012 Atmospheric CO₂ over the last 1000 years: A high-resolution record from the West Antarctic Ice Sheet (WAIS) Divide ice core *Glob. Biogeochem. Cycl.* **26** GB2027
- Ammann C M, Joos F, Schimel D S, Otto-Bliesner B L and Tomas R A 2007 Solar influence on climate during the past millennium: Results from transient simulations with the NCAR Climate System Model *Proc. Natl. Acad. Sci USA* **104** 3713-3718
- Benjamini Y and Hochberg Y 1995 A Practical and Powerful Approach to Multiple Testing *J. Roy. Stat. Soc. B* **57** 289-300
- Böhm R, Auer I, Schöner W, Ganekind M, Gruber C, Jurkovic A and Orlik M and Ungersböck M 2009 *Eine neue Webseite mit instrumentellen Qualitäts-Klimadaten für den Grossraum Alpen zurück bis 1760*. Wiener Mitteilungen **216** Hochwässer: Bemessung, Risikoanalyse und Vorhersage
- Böhm R, Jones P D, Hiebl J, Brunetti M, Frank D C and Maugeri M 2010 The early instrumental warm-bias: a solution for long Central European temperature series 1760–2007 *Clim. Change* **101** 41-67
- Braconnot P, Harrison S P, Kageyama M, Bartlein P J, Masson-Delmotte V, Abe-Ouchi A, Otto-Bliesner B L and Zhao Y 2012 Evaluation of climate models using paleoclimate data *Nat. Clim. Change* **2** 417–424
- Briffa K R, Jones P D, Bartholin T S, Eckstein D, Schweingruber F H, Karlén W, Zetterberg P and Eronen M 1992 Fennoscandian summers from AD 500: temperature changes on short and long timescales *Clim. Dyn.* **7** 111–119
- Büntgen U, Frank D C, Nievergelt D, Esper J 2006 Summer temperature variations in the European Alps, AD 755-2004 *J. Clim.* **19** 5606-5623
- Büntgen U, Franke J, Frank D C, Wilson R, González-Rouco J F and Esper J 2010 Tree-ring proxy number and location to reconstruct European climate variability *Clim. Res.* **41** 125-130
- Büntgen U, Tegel W, Nicolussi K, McCormick M, Frank D, Trouet V, Kaplan J, Herzig F, Heussner U, Wanner H, Luterbacher J and Esper J 2011 2500 years of European climate variability and human susceptibility *Science* **331** 578-582
- Büntgen U, Frank D C, Neuenschwander T and Esper J 2012 Fading temperature sensitivity of Alpine tree grow at its Mediterranean margin and associated effects on large-scale climate reconstructions *Clim. Change* **114** 651-666
- Büntgen U, Kyncl T, Ginzler C, Jacks D S, Esper J, Tegel W, Heussner K U and Kyncl J 2013 Filling the Eastern European gap in millennium-long temperature reconstructions *Proc. Natl. Acad. Sci. USA* **110** 1773–1778
- Bürger G and Cubasch U 2006 Climate reconstruction by regression – 32 variations on a theme *Tellus A* **58** 227–235
- Camuffo D 2002 History of the long series of the air temperature in Padova (1725-today) *Clim. Change* **53** 7-76
- Camuffo D and Bertolin C 2012a The earliest temperature observations in the world: the Medici Network (1654–1670) *Clim. Change* **111** 335-363
- Camuffo D and Bertolin C 2012b The earliest spirit-in-glass thermometer and a comparison between the earliest CET and Italian observations *Weather* **67** 206-209

- Camuffo D et al. 2010 500-year temperature reconstruction in the Mediterranean Basin by means of documentary data and instrumental observations. *Clim. Change* **101** 169-199
- Christiansen B, Schmith T and Thejll P 2009 A surrogate ensemble study of climate reconstruction methods: stochasticity and robustness *J. Clim.* **22** 951–976
- Christiansen B 2011 Reconstructing the NH Mean Temperature: Can Underestimation of Trends and Variability be avoided? *J. Clim.* **24** 674–692
- Cook E R, Briffa K R and Jones P D 1994 Spatial regression methods in dendroclimatology: A review and comparison of two techniques *Int. J. Climatol.* **14** 379–402
- Cook E R, Briffa K R, Meko D M, Graybill D A and Funkhouser G 1995 The “segment length curse” in long tree-ring chronology development for palaeoclimatic studies *The Holocene* **5** 229-237
- Crowley T J and Unterman M B 2013 Technical details concerning development of a 1200–year proxy index for global volcanism *Earth Syst. Sci. Data* **5** 187–197
- Diaz H D and Wahl E R 2015 Recent California Water Year Precipitation Deficits: A 440-year Perspective *J. Clim.* **28** 4637-4652
- Dobrovolný P et al. 2010 Temperature reconstruction of Central Europe derived from documentary evidence since AD 1500 *Clim. Change* **101** 69-107
- Dorado Liñán I, Büntgen U, González-Rouco F J, Zorita E, Montávez J P, Gómez-Navarro J J, Brunet M, Heinrich I, Helle G and Gutiérrez E 2012 Estimating 750 years of temperature variations and uncertainties in the Pyrenees by tree-ring reconstructions and climate simulations *Clim. Past* **8** 919-933
- Dunn O J 1961 Multiple Comparisons Among Means *J. Amer. Stat. Assoc.* **56** 52–64
- Düthorn E, Holzkämper S, Timonen M and Esper J 2013 Influence of micro-site conditions on tree-ring climate signals and trends in central and northern Sweden *Trees* **27** 1395–1404
- Düthorn E, Schneider L, Konter O, Schön P, Timonen M and Esper J. 2015 On the hidden significance of differing micro-sites on tree-ring based climate reconstructions *Silva Fennica* **49** doi.org/10.14214/sf.1220
- Esper J, Cook E and Schweingruber F H 2002 Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability *Science* **295** 2250–2253
- Esper J, Cook E R, Krusic P J, Peters K, Schweingruber F H 2003 Tests of the RCS method for preserving low-frequency variability in long tree-ring chronologies *Tree-Ring Res.* **59** 81-98
- Esper J, Frank, D C, Timonen M, Zorita E, Wilson R J S, Luterbacher J, Holzkämper S, Fischer N, Wagner S, Nievergelt D, Verstege A and Büntgen U 2012 Orbital forcing of tree-ring data *Nat. Clim. Change* **2** 862–866
- Esper J, Düthorn E, Krusic P, Timonen M, and Büntgen U 2014 Northern European summer temperature variations over the Common Era from integrated tree-ring density records. *J. Quat Sci.* **29** 487-494
- Fernández-Donado L, González-Rouco J F, Raible C C, Ammann C M, Barriopedro D, García-Bustamante E, Jungclaus J H, Lorenz S J, Luterbacher J, Phipps S J, Servonnat J, Swingedouw D, Tett S F B, Wagner S, Yiou P and Zorita E 2013 Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium *Clim. Past* **9** 393-421

- Field A 2013 *Discovering statistics using IBM SPSS statistics* Sage 952pp
- Frank D, Büntgen U, Böhm R, Maugeri M, Esper J 2007 Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at a moving target *Quat. Sci. Rev.* **26** 3298-3310
- Gao C, Robock A and Ammann C 2008 Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models *J. Geophys. Res.* **113** D23111
- Gao C, Robock A and Ammann C 2012 Correction to “Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models” *J. Geophys. Res.* **117** D16112
- Gelman A et al. 2003 *Bayesian Data Analysis*, Chapman Hall CRC
- González-Rouco J F, Beltrami H, Zorita E and von Storch H 2006 Simulation and inversion of borehole temperature profiles in surrogate climates: Spatial distribution and surface coupling *Geophys. Res. Lett.* **33** L01703
- Guiot J, Corona C and ESCARSEL members 2010 Growing Season Temperatures in Europe and Climate Forcings Over the Past 1400 Years *PLoS ONE* **5** e9972
- Gunnarson B E, Linderholm H W and Moberg A 2011 Improving a tree-ring reconstruction from west-central Scandinavia – 900 years of warm-season temperatures *Clim. Dyn.* **36** 97-108
- Hegerl G, Luterbacher J, González-Rouco F J, Tett S F B, Crowley T and Xoplaki E 2011 Influence of human and natural forcing on European seasonal temperatures *Nat. Geosci.* **4** 99-103
- Hofer D, Raible C C and Stocker T F 2011 Variations of the Atlantic meridional overturning circulation in control and transient simulations of the last millennium *Clim. Past.* **7** 133-150
- Jones P D, Osborn T J and Briffa K R 1997 Estimating Sampling Errors in Large-Scale Temperature Averages *J. Clim.* **10** 2548–2568
- Jones P D, Lister D H, Osborn T J, Harpham C, Salmon M and Morice C P 2012 Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010 *J. Geophys. Res.* **117** D05127
- Jungclauss J H et al. 2010 Climate and carbon-cycle variability over the last millennium *Clim. Past* **6** 723-737
- Jungclauss J H, Lohmann K and Zanchettin D 2014 Enhanced 20th century heat transfer to the Arctic simulated in the context of climate variations over the last millennium. *Clim. Past* **10** 2201-2213
- Jurin J 1723 *Invitatio ad observationes Meteorologicas communi consilio instituendas a Jacobo Jurin M. D. Soc. Reg. Secr. et Colleg. Med. Lond. Socio Philosophical Transactions* **379** 422–427
- Klingbjær P and Moberg A 2003 A composite monthly temperature record from Tornedalen in northern Sweden 1802–2002 *Inter. J. Climatol.* **23** 1465–1494
- Landrum L, Otto-Bliesner B L, Conley A, Lawrence P, Rosenbloom N and Teng H 2013 Last Millennium Climate and Its Variability in CCSM4 *J. Clim.* **26** 1085–1111
- Larocque-Tobler I, Grosjean M, Heiri O, Trachsel M and Kamenik C 2010 Thousand years of climate change reconstructed from chironomid subfossils preserved in varved lake Silvaplana, Engadine, Switzerland *Quat. Sci. Rev.* **29** 1940-1949

- Lean J, Wang Y and Sheeley Jr N 2002 The effect of increasing solar activity on the Sun's total and open magnetic flux during multiple cycles- Implications for solar forcing of climate *Geophys. Res. Lett.* **29**, 2224
- Linderholm HW, Zhang P, Gunnarson BE, Björklund J, Farahat E, Fuentes M, Rocha E, Salo R, Seftigen K and Stridbeck P 2014 Growth dynamics of tree-line and lake-shore Scots pine (*Pinus sylvestris* L.) in the central Scandinavian Mountains during the Medieval Climate Anomaly and the early Little Ice Age *Front. Ecol. Evol.* **2** doi.10.3389/fevo.2014.00020
- Ljungqvist F C, Krusic P J, Brattström G and Sundqvist H S 2012 Northern Hemisphere temperature patterns in the last 12 centuries *Clim. Past* **8** 227-249
- Luoto T P, Sarmaja-Korjonen K, Nevalainen L and Kauppila T 2009 A 700 year record of temperature and nutrient changes in a small eutrophied lake in southern Finland *The Holocene* **19** 1063-1072
- Luoto T P and Helama S 2010 Palaeoclimatological and palaeolimnological records from fossil midges and tree-rings: the role of the North Atlantic Oscillation in eastern Finland through the Medieval Climate Anomaly and Little Ice Age *Quat. Sci. Rev.* **29** 2411–2423
- Luterbacher J, Dietrich D, Xoplaki E., Grosjean M and Wanner H 2004 European seasonal and annual temperature variability, trends, and extremes since 1500 *Science* **303** 1499-1503
- Manley G 1974 Central England temperatures: monthly means 1659 to 1973 *Quart. J. R. Meteorol. Soc.* **100** 389–405
- Mann M E, Zhang H, Hughes M K, Bradley R S, Miller S K, Rutherford S and Ni F 2008 Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia *Proc. Natl. Acad. Sci. USA* **105** 13252-13257
- Martín-Chivelet J, Muñoz-García M B, Edwards R L, Turrero M J and Ortega A I 2011 Land surface temperature changes in Northern Iberia since 4000 yr BP, based on $\delta^{13}\text{C}$ of speleothems *Glob. Plan. Change* **77** 1-12
- Masson-Delmotte V et al 2013 Information from Paleoclimate Archives. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T. F., D. Qin, G-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA
- Melvin T M, Grudd H and Briffa K R 2013 Potential bias in 'updating' tree-ring chronologies using regional curve standardisation: Re-processing 1500 years of Torneträsk density and ring-width data *The Holocene* **23** 364-373
- Middleton W E K 1966 *A history of the thermometer and its use in meteorology*. Johns Hopkins Press, Baltimore
- Moberg A, Bergström H, Ruiz Krigsman J, Svanered O 2002 Daily air temperature and pressure series for Stockholm (1756-1998) *Clim. Change* **53** 171-212
- Moberg A and Bergström H 1997 Homogenization of Swedish temperature data. Part III: The long temperature records from Stockholm and Uppsala *Int. J. Climatol.* **17** 667-699
- Moberg A, Alexandersson H, Bergström H and Jones P D 2003 Were southern Swedish summer temperatures before 1860 as warm as measured? *Int. J. Climatol.* **23** 1495-1521
- Nordli P Ø, Lie Ø, Nesje A and Dahl S O 2003 Spring-summer temperature reconstruction in western Norway 1734–2003: a data-synthesis approach *Int. J. Clim.* **23** 1821–1841

- PAGES 2k Consortium 2013 Continental-scale temperature variability during the last two millennia *Nat. Geosci.* **6** 339-346
- PAGES2k-PMIP3 group 2015 Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium *Clim. Past Discuss.* **11** 2483–2555
- Parker D E, Legg T P and Folland C 1992 A new daily Central England temperature series 1772–1991 *Int. J. Climatol.* **12** 317–342
- Phipps S, Rotstayn L, Gordon H, Roberts J, Hirst A and Budd W 2011 The CSIRO Mk3L climate system model version 1.0 Part 1: Description and evaluation *Geosci. Mod. Develop.* **4** 483-509
- Phipps S J, Gergis J, McGregor H V, Gallant A J E, Neukom R, Stevenson S, van Ommen T D, Brown J R, Fischer M J and Ackerley D 2013 Palaeoclimate data-model comparison: Concepts and application to the climate of Australasia over the past 1500 years *J. Clim.* **26** 6915–6936
- Popa I and Kern Z 2009 Long-term summer temperature reconstruction inferred from tree-ring records from the Eastern Carpathians *Clim. Dyn.* **32** 1107-1117
- Pribyl K, Cornes R C and Pfister C 2012 Reconstructing medieval April-July mean temperatures in East Anglia, 1256–1431 *Clim. Change* **113** 393-412
- Risken H 1989 *The Fokker-Planck Equation*, Springer-Verlag Berlin
- Schmidt G A et al. 2011 Climate forcing reconstructions for use in PMIP simulations of the last millennium *Geosci. Mod. Dev.* **4** 33-45
- Schmidt G A et al 2012 Climate forcing reconstructions for use in PMIP simulations of the last millennium *Geosci. Mod. Dev.* **5** 185-191
- Schmidt G A et al. 2014 Using palaeo-climate comparisons to constrain future projections in CMIP5 *Clim. Past* **10** 221–250
- Schneider T 2001 Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values *J. Clim.* **14** 853-871
- Schurer A, Hegerl G C, Mann M E, Tett S F B and Phipps S 2013 Separating forced from chaotic variability over the last millennium *J. Clim.* **26** 6954–6973
- Seim A, Büntgen U, Fonti P, Haska H, Herzig F, Tegel W, Trouet V and Treydte K 2012 The paleoclimatic potential of a millennium-long tree-ring width chronology from Albania *Clim. Res.* **51** 217-228
- Servonnat J, Yiou P, Khodri M, Swingedouw D and Denvil S 2010 Influence of solar variability, CO₂ and orbital forcing between 1000 and 1850 AD in the IPSLCM4 model *Clim. Past* **6** 445-460
- Sleptsov A M and Klimenko V V 2003 Complex Reconstruction of Climate of East Europe over the Last 2,000 years *Proc. Russ. Geogr. Soc.* **6** 45–53 (in Russian)
- Smerdon J E and Kaplan A 2007 Comments on "Testing the fidelity of methods used in proxy-based reconstructions of past climate": The role of the standardization interval *J. Clim.* **20** 5666–5670
- Smerdon J E 2012 Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments *WIREs Clim. Change* **3** 63-77

- Sueyoshi T, Ohgaito R, Yamamoto A, Chikamoto M O, Hajima T, Okajima H, Yoshimori M, Abe M, O'ishi R, Saito F, Watanabe S, Kawamiya M and Abe-Ouchi A 2013 Set-up of the PMIP3 paleoclimate experiments conducted using an Earth system model, MIROC-ESM *Geosci. Model Dev.* **6** 819-836
- Stemler T, Werner J P, Benner H and Just W 2007 Stochastic Modeling of Experimental Chaotic Time Series *Phys. Rev. Lett.* **98** 044102
- Swingedouw D, Terray L, Cassou C, Voldoire A, Salas-Mélie D and Servonnet J 2010 Natural forcing of climate during the last millennium: fingerprint of solar variability *Clim. Dyn.* **36** 1349-1364
- Taylor K E, Stouffer R J and Meehl G A 2012 An Overview of CMIP5 and the experiment design *Bull. Amer. Meteor. Soc.* **93** 485-498
- Tegel W, Vanmoerkerke J and Büntgen U 2010 Updating historical tree-ring records for climate reconstruction *Quat. Sci. Rev.* **29** 1957-1959
- Tett S F B, Betts R, Crowley T J, Gregory J, Johns T C, Jones A, Osborn T J, Ostrom E, Roberts D L and Woodgate M J 2007 The impact of natural and anthropogenic forcings on climate and hydrology since 1550 *Clim. Dyn.* **28** 3-34
- Tingley M P and Huybers P 2010a A Bayesian algorithm for reconstructing climate anomalies in space and time. Part 1. Development and applications to paleoclimate reconstructions problems *J. Clim.* **23** 2759–2781
- Tingley M P and Huybers P 2010b A Bayesian algorithm for reconstructing climate anomalies in space and time. Part 2. Comparison with the Regularized Expectation-Maximization Algorithm *J. Clim.* **23** 2782–2800
- Tingley M and Huybers P 2013 Recent temperature extremes at high northern latitudes unprecedented in the past 600 years *Nature* **496** 201-205
- Trachsel M et al. 2012 Multi-archive summer temperature reconstruction for the European Alps, AD 1053-1996 *Quat. Sci. Rev.* **46** 66-79
- van Engelen A F V and Nellestin J W 1996 Monthly, seasonal, and annual means of the air temperature in tenths of centigrades in De Bilt, Netherlands, 1706-1995. (KNMI Report, Climatological Services Division, 1996), pp 6
- von Storch H, Zorita E, Jones J M, Dimitriev Y, González-Rouco J F and Tett, S F B 2004: Reconstructing past climate from noisy data *Science* **306** 679–882
- Wagner S, Widmann M, Jones J, Haberzettl T, Lücke A, Mayr C, Ohlendorf C, Schäbitz F and Zolitschka B 2007 Transient simulations, empirical reconstructions and forcing mechanisms for the Mid-Holocene hydrological climate in Southern Patagonia *Clim. Dyn.* **29** 333-355
- Wahl E R and Smerdon J E 2012 Comparative performance of paleoclimatic field and index reconstructions derived from climate proxies and noise-only predictors *Geophys. Res. Lett.* **39** L06703
- Werner J P, Luterbacher J and Smerdon J E 2013 A Pseudoproxy Evaluation of Bayesian Hierarchical Modelling and Canonical Correlation Analysis for Climate Field Reconstructions over Europe *J. Clim.* **26** 851-867
- Werner J P, Toreti A and Luterbacher J 2014 Stochastic models for climate reconstructions -- how wrong is too wrong? *Nolta Proc.* **24** 528-531
- Wu T W A 2012 Mass-Flux Cumulus Parameterization Scheme for Large-scale Models, description and Test with Observations *Clim. Dyn.* **38** 725-744

Zhang H, Yuan N, Xoplaki E, Werner J P, Büntgen U, Esper J, Treydte K and Luterbacher J 2015 Modified climate with long term memory in tree ring proxies *Environ. Res. Lett.* **10** 084020 doi:10.1088/1748-9326/10/8/084020

Zhou T, Li B, Man W, Zhang L and Zhang J 2011 A Comparison of the Medieval Warm Period, Little Ice Age and 20th Century Warming simulated by the FGOALS Climate System Model *Chin. Sci. Bull.* **56** 3028-304