nature geoscience

SUPPLEMENTARY INFORMATION

DOI: 10.1038/NGEO1057

Influence of human and natural forcing on European seasonal temperatures

Gabriele Hegerl[1], Juerg Luterbacher[2], Fidel González-Rouco[3], Simon Tett[1], Thomas Crowley[1] and Elena Xoplaki[4]

Proxy reconstruction

We used surface air temperature reconstructions over land (25°W to 40°E and 35°N to 70°N) given on a 0.5°x0.5° grid [1,2,3], which are based on CRU TS2.1[4] for statistical calibration. The reconstruction is based on a field reconstruction method, with a calibration period of 1901-1960 and a verification period of 1961-1995. The reconstructions are seasonally resolved from 1500 to 1658, when only proxy data (tree rings, ice cores, documentary-based temperature indices, sea ice information) are available, and monthly thereafter. For each set of proxy data available at any given time a statistical model is fitted (yielding about 500 different statistical models for different data availability). Early instrumental data are also used in the reconstruction approach after 1659. After approximately 1750, the main information in these reconstructions comes from instrumental data. The reconstructions were used until 1900, after that they were substituted with observation-based data[4]. In figures 1, SI1 and SI2, European averages of the most recent instrumental records from CRUTEM3v[5] are used for comparison, small differences to the series shown in figure 1 in the main paper likely illustrate uncertainties in gridding (5°x5° grid rather than the 0.5x°x0.5° higher resolution grid the reconstruction is calibrated to) and input data available for the final gridded datasets. We used a low-resolution version of the reconstruction for the field comparisons and timeseries attribution analysis, which has been transformed to the ECHO-G grid.

We have tested the robustness of our fingerprint results to substituting the newest instrumental data[5] for the European dataset for 1900-1996. Both external forcing and individual EBM fingerprint detection results do not change if this revised reconstruction is used. The only exception is that autumn results, which are just <5% significant (table 1, body of paper), now just that threshold, yielding them 10% significant.

In the data-poor early period, the field reconstruction technique may tend to decrease variance[6,7]. However, during the period when instrumental data are available, the reconstruction at the appropriate gridpoint replicates long-term instrumental datasets well (for example, the Central England Temperature series, Schurer, pers. com. 2010). Loss of early variance would make it more difficult to detect the response to external forcing relative to the estimate of internal variability (which is based on the models that show a consistent variance over time and hence overestimate reconstructed variance). Therefore, spurious detection of response to external forcing based on the less reliable earlier reconstruction period is unlikely. Uncertainties in the reconstruction are discussed in the original papers[1-3], and are smaller for this reconstruction using long instrumental series than they would be for proxy-only based reconstructions due to the strong constraints from the instrumental data, and the large spatial autocorrelation of annual and decadal temperature changes. However, prior to the instrumental data, uncertainties in the extent to which for example, treering data resolve the full variance in the past, can affect the overall level of variability, and the reconstruction technique tends to bias the variance low in the case of noisy information. Uncertainty in spatial detail of the reconstruction varies between locations and season, results are most robust close to instrumental points and least robust early in the reconstruction or in areas where no proxy information is available. Therefore, we focus on European mean temperatures for identification of the effect of external forcings, which is fairly well constrained by instrumental data, particularly, after the 17th century.

DOI: 10.1038/NGEO1057

Both temperature reconstructions as well as reconstructions of past forcing used to drive the climate models are associated with uncertainty, which is, however, independent from each other. Therefore, agreement between models and reconstructions is unlikely to occur due to errors in either dataset.

Climate model simulations

ECHO-G

ECHO-G is a coupled atmosphere-ocean general circulation model (AOGCM), consisting of the ECHAM4 atmospheric general circulation model and the Hamburg Ocean Primitive Equation model HOPE-G. Both components were developed at the Max-Planck Institute for Meteorology in Hamburg and coupled with the OASIS-Software. The ECHO-G model has been evaluated against instrumental data in several studies (e.g., ref. 8).

The atmospheric component ECHAM4 has a horizontal resolution of T30 (approx. 3.75° x 3.75° longitude / latitude) and 19 vertical levels, five of them located above 200hPa, the highest being 10hPa, thus having a rather coarse resolution of the stratosphere. The oceanic component HOPE-G has a T42 resolution (approx. 2.8° x 2.8° longitude x latitude) with a decrease in meridional grid-point separation towards the equator to a value of 0.5°. Due to the interactive coupling between ocean and atmosphere, ECHO-G needs annual mean flux adjustments to avoid a significant climate drift. Thus additional fluxes of heat and freshwater are applied to the ocean. This flux adjustment is constant in time and its global integral vanishes.

This comparison uses a 1000-yr simulation of the last millennium with the ECHO-G model[9] that incorporates changes in the values of some natural and anthropogenic forcing factors based on reconstructions provided by ref. 10: solar irradiance, the radiative effects of stratospheric volcanic aerosols and the concentrations of some greenhouse gases (CO_2, CH_4 and N_2O). Volcanic aerosols are included as equivalent reduction of solar constant values and therefore do not discriminate for timing and location of eruption. Tropospheric sulphate aerosols and changes in vegetation as well as orbital changes are not included in this simulation. A more detailed description of the forcing and type of model setup can be found in refs 11 and 12. Despite the lack of tropospheric aerosols in this model simulation, it is quite similar in correlations over the millennium and the post-1675 period to the other simulations (see table SI1). EPOCH analysis results from ECHO-G alone (not shown) show very similar response to volcanic eruptions as for the other two models, although the other two models directly implement volcanic aerosol in the stratosphere, which leads to stratospheric warming through absorption of infra-red radiation via the aerosol layer in both other models.

HadCM3

We use a simulation of the last five centuries with HadCM3[13]. HadCM3 does not require flux adjustment to prevent large climate drifts in the model[14]. The atmospheric component HadAM3 is a version of the United Kingdom Meteorological Office (UKMO) unified forecast and climate model with a horizontal grid spacing of 2.5° (latitude) x 3.75° (longitude) (about T42) and 19 vertical levels. The ocean component has 20 levels with a spatial resolution of 1.25° x 1.25° degrees.

In this study, two HadCM3 runs were merged: A simulation driven with natural forcings from 1500 to 1749 has been merged with a simulation driven with natural and anthropogenic forcings spanning the period 1750-1999. The natural forcings run is driven by prescribed changes in volcanic forcing (using an updated version of the Crowley volcanic forcing[10] and

implemented as latitudinally varying stratospheric aerosol forcing) and orbital forcing, while anthropogenic forcing factors were fixed at estimated pre-industrial values. The simulation driven with all forcings combined is driven by prescribed changes in volcanic forcing, solar irradiance, orbital forcing, greenhouse gases, tropospheric sulphate aerosol, stratospheric ozone and land-use/land-cover (for details see ref 13). The simulation used here is downloaded from the SOAP project (http://www.cru.uea.ac.uk/cru/projects/soap/data/model/), and uses unrealistically large ozone forcing from 1970 on. However, the difference to a simulation where the last 3 decades have been rerun, using more realistic ozone forcing, is small for European seasonal mean temperatures, and consistent with internal variability. The corrected version is used for splicing with the future simulation for figure SI2. The HadCM3 simulation was extended to 2100 using the A2 scenario with greenhouse gas and other drivers[14].

NCAR CSM 1.4 Model

A simulation of the last millennium has also been performed with the NCAR CSM1.4[15]. CSM1.4 is a global coupled ocean-atmosphere sea ice and land surface climate model with approximately 3.75◦ × 3.75◦ resolution. The simulation has been run over the period 850 to 1999 AD, and forced with reconstructed time histories of solar irradiance, aerosol loadings from explosive volcanism[16], greenhouse gases and anthropogenic sulfate aerosols. Orbital parameters and land use changes are not included as forcings. Any potential long-term drift is removed.

Multimodel technique

All model data have been transformed to the ECHO-G grid for comparison.
There are some differences in the external forcings used in the model simulations. For example, ECHO-G implements volcanic forcing through reduction in solar radiation, while both other models apply sulfate aerosol loadings in the stratosphere directly. Also, the solar forcing used is larger in the ECHO-G simulation than in both other simulations. Furthermore, only HadCM3 forcing includes changes in orbital forcing and land cover change. Therefore, differences between the model simulations reflect a combination of different realizations of internal variability, differences in the response to the same external forcing, and differences in the external forcing. The solar forcing is perhaps the most uncertain forcing. Though there is considerable discussion[17-19] about the amplitude of changes in past Total Solar Irradiance (TSI), current estimates seem to support changes in the order of 0.1% of TSI between the late Maunder Minimum and the late 20th century. The CSM simulation used herein incorporates a 0.1 % increase in TSI for this period while the HadCM3 and ECHO-G simulations use 0.25% and 0.3%, respectively.
Despite these uncertainties, the relatively large variability in external forcing options used in the various models yields a good first guess of the role of model and forcing uncertainty in assessments of our ability to simulate past climate change.

Role of total forcing on reconstructions

Comparison of simulated and reconstructed mean temperatures

For this comparison, a low-resolution version of the reconstruction using the ECHO-G grid and landmask has been used, and all model data have been transformed to the same grid and landmask. A comparison of variability in models and reconstruction in the past as well as

DOI: 10.1038/NGEO1057

correlation studies illustrate to what extent the magnitude and role of forced variability varies between the seasons (Fig. SI1).

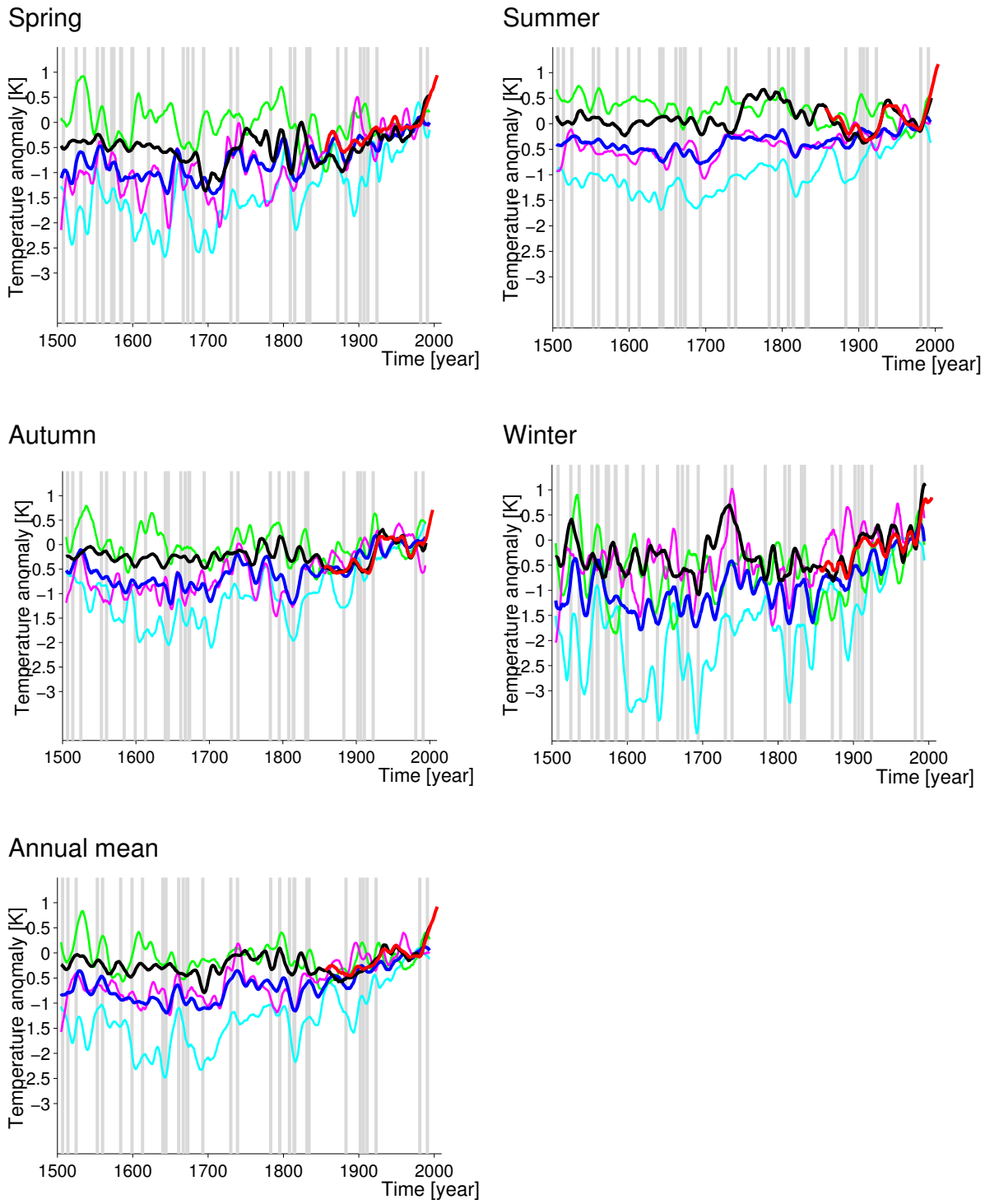

Figure SI1: *As figure 1, but for European mean temperature [K] in all seasons and annual average. Reconstruction (black) compared to climate model simulations (green: HadCM3; magenta: CCSM; cyan: ECHO-G; dark blue and bold: multimodel mean) in spring (top left), summer (top right), autumn (center left), winter (center right) and annual mean data (bottom). All data have been smoothed by a filter removing timescales below 15 years (by applying a running 11- and 7- year filter consecutively). As in figure 1, body of paper, the time of volcanic eruptions has been marked by grey bars. Recent instrumental data[5] are shown in red for comparison.*

DOI: 10.1038/NGEO1057

Table SI1: *Extension of table 1 in main text: Correlations of decadally averaged (running 11-yr mean) reconstructed European mean temperatures with those simulated by different individual climate model simulations (row 1-3), the average of all three simulations ('Multimodel', row 4), and the average correlation between an individual simulation and the average of both others ('Average Intermodel corr'); in brackets for the period < 1900). The bottom two rows give correlations between the reconstruction and the multimodel mean over two periods, namely the better reconstructed more recent period, and over the pre-instrumental period. Significance of correlations is not assessed, as the fingerprint detection results provide more rigorous assessments for the role of forcings.*

	annual	spring	summer	autumn	winter
ECHO-G	0.5	0.42	0.17	0.39	0.38
HadCM3	0.41	0.15	0.05	0.24	0.26
CCSM	0.35	0.37	0.13	0.18	0.41
Multimodel	0.59	0.50	0.19	0.38	0.51
Average Intermodel Corr.	0.37 [0.10]	0.24 [0.03]	0.17 [0.16]	0.35 [0.19]	0.28 [-0.01]
Multimodel 1675 on	0.60	0.60	0.15	0.37	0.55
Multimodel 1675-1900	0.35	0.30	0.41	0.08	0.30

Table SI2: *Standard deviation of annual and seasonal mean instrumental and model simulated data. Annual mean data are given in K, seasonal values are expressed relative to the standard deviation of annual mean values. Note that in all models and reconstruction, variability in all season but summer exceeds that of annual means, most strongly in winter. The ratio of variance between seasonal and annual means in the models, the reconstruction and instrumental data is reasonably similar.*

	annual [K]	spring/annual	summer/ annual	autumn/ annual	winter/annual
ECHO-G	0.89	1.45	0.65	1.10	2.05
HadCM3	0.63	1.42	0.98	1.38	2.27
CCSM	0.79	1.75	0.68	1.34	2.24
reconstructed	0.47	1.45	0.95	1.12	2.54
Instrumental	0.51	1.36	1.00	1.20	2.07

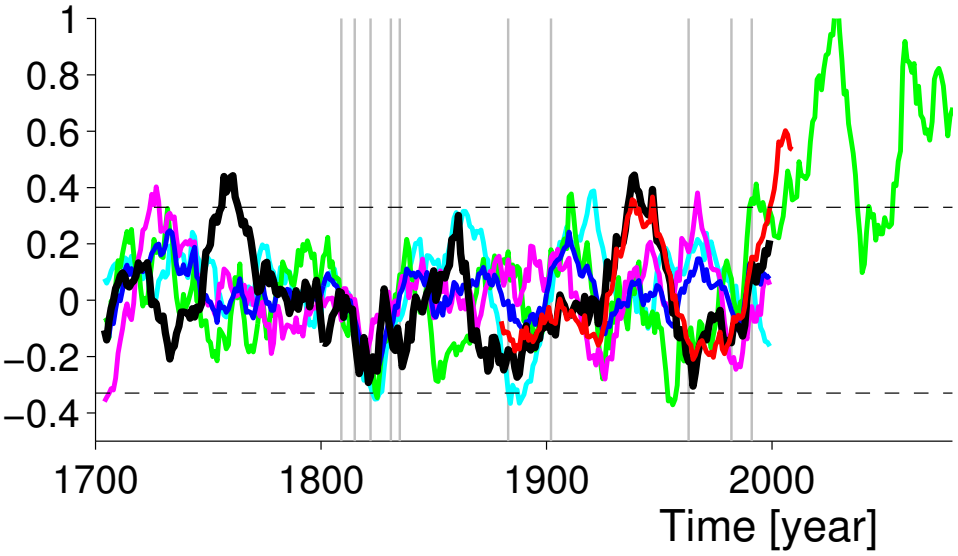

Figure SI2: ***Significance of recent summer trends***. *Comparison of running summer (JJA) 30-year trends (values plotted against the end of the respective 30 year period; in K/Decade) from proxy reconstruction from the trend 1675-1704 on (black), instrumental data (red, ending with the trend 1980 to 2009) and the climate model simulations in blue (multi model mean), cyan (ECHO-G), magenta (CCSM) and green (HadCM3); data plotted against the last year of the trend; in K/Decade (note that reconstructed trends prior to 1675 may have a low bias, see text, and are not used here). The black dashed lines indicate two standard deviations for 30-year trends from the reconstructions (2-sigma from individual model runs is slightly smaller, not shown). The HadCM3 simulation was extended to 2100 using the A2 scenario with greenhouse gas and other drivers*[14].

While for most seasons, recent multi-decadal trends are not very unusual compared to the past in both model and reconstruction, the recent observed increase in summer temperature (figure SI2) is highly unusual relative to trends earlier in the proxy period, yielding all trends from the period 1971-2000 to the final trend 1980-2009 significant at the <5% one-sided level (using conservative estimates of the number of degrees of freedom in the estimate). The present rapid increase in summer temperatures is consistent with the expected response to anthropogenic forcing, as evident from the continued model simulation[14].

Fingerprint method for spatial averages of seasonal mean temperature

The spatial average of European seasonal temperatures has been analyzed by regressing them both onto fingerprints of all forcings combined; and by simultaneously regressing them onto fingerprints for combined anthropogenic (greenhouse gas and aerosols), solar and volcanic forcing.

Fingerprints

While the fingerprint method is explained in the 'methods' section (see also ref. 20), several technical details are worth mentioning here:

1. Instead of an analytic formation of the uncertainty in β total least square (TLS) fingerprint estimates as in ref 20, we used random variability superimposed on the observations, which has been tested to yield similar results (P. Stott, pers. Com., 2009). We find that in some seasons, the model fingerprint has to be substantially reduced in amplitude to match our

DOI: 10.1038/NGEO1057

reconstruction, despite application of a TLS fit to account for noise in the fingerprint (table 1). However, it is unclear if this reduced variance in the reconstructed temperature signal is significant given uncertainties in forcings and reconstruction.

2. Errors in the overall magnitude of forcing or climate model response will not affect detection results[21]. Note that we do not 'optimize' fingerprints in order to improve signal-to-noise ratios as often done in analysis of instrumental records. The reason is that benefit from optimization would particularly arise from spatial patterns of the fingerprint, which are less reliable than for instrumental data (as discussed in body of paper).

Noise estimate

The noise estimate is crucial to assess if an observed change is significant, and to estimate uncertainty ranges for response patterns, and regression coefficients. We used two different data sources as noise estimate:

1. The climate model simulations all contain internal climate variability in addition to the response to external forcing. Usually, this internal variability is isolated by removing the common forced signal from model simulations. However, due to the differences in forcing in the simulations, removing the multimodel average resulted in very strong secular trends remaining in some simulations (as can be deduced from Figure SI1). Therefore, the combined response to all forcings in the EBM has been fitted to each individual model simulation and removed from each simulation separately (see figure SI2). This method will result in a less than perfect removal of the forced component, as the forcings will not be the same as in the EBM for the model simulations. However, residual responses to forcing in the variability estimate will make detection of external influences more difficult and thus yield conservative attribution results (i.e., it is much more likely that a significant result is not detected than falsely detected). The residuals from all three simulations have been concatenated and shifted by varying numbers of years to yield a large ensemble of noise simulations and thus samples of variability over the time period covered (samples starting after year 1000 are augmented by the beginning of the noise timeseries). Generally, it is more difficult to detect the response to forcing against this model derived estimate of internal variability, as it is larger than the residual from the reconstruction (see also table SI1).We use this model-based noise estimate for the main results in the paper. An alternative to the approach taken here would have been to use control simulations of the respective model, which we did not do for reasons of simplicity (but which should make no difference unless signal and variability interact nonlinearly, which appears generally not be the case, see discussion in ref. 3, main paper).
2. The residual variability in the reconstruction after regressing out the response to all forcings. This is similar as in ref. 7 , and the unexplained variability in the regression (eqn 1, 2, methods section) is used as noise estimate. However, the number of effective degrees of freedom in these samples is very limited due to autocorrelation even if residuals are shifted by varying increments (we determined this effect by calculating the autocorrelation and thus the decorrelation time of 20-yr offset residuals; the effective sample size was then estimated from the number of samples separated by the decorrelation time, increased by a factor of 1.5 to account for better sampling in running estimates[20], and reduced by the number of signals fitted to the data). Furthermore, the possible reduction of variance in the earlier, data sparse period of the reconstruction may lead to temporal inhomogeneity and to an underestimate of variability when using data from earlier in the reconstruction.

The model-derived estimate has more variance than the residual, suggesting that the use of the climate model internal variability is conservative. The reason for the difference could be both a slight loss of variance in the reconstruction for the earlier, data sparse period; overestimated

variability in the models, or/and additional variability originating from incomplete removal of the forced component from the climate models.

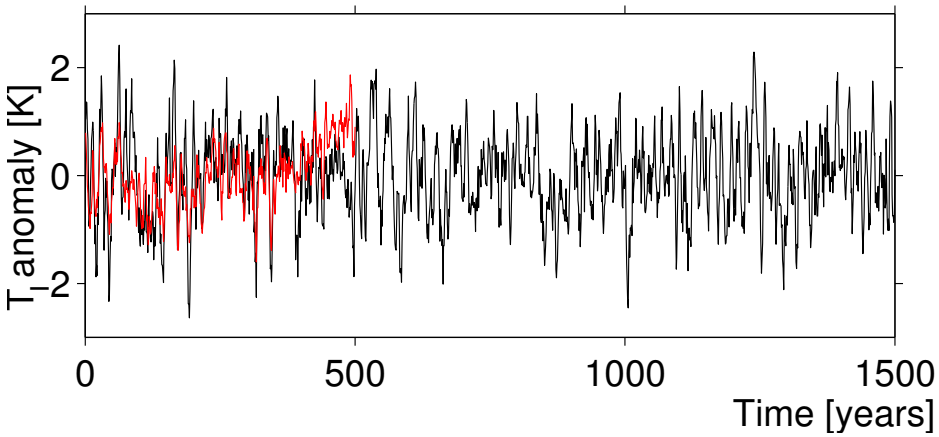

Figure SI3: ***Estimate of internal variability (winter)***. *Variability in the three coupled model integrations after an Energy Balance Model response to all forcings was fitted to each individual simulation and subtracted (black) compared to the fingerprint of all forcings combined from the multimodel mean (red). From this 1500 yr timeseries, a large sample of 500-yr timeseries can be selected as noise estimates starting at different points in time. Most of these will not be aligned with any residual signal in the simulations (exceptions are chunk selected starts in years 1, 501 or 1001).*

Results

Table SI3: *Detection results for the fingerprint of external forcings in seasonal European mean temperature data 1500-2000, expanding on table 1 of the main text. All results are based on 5-yr smoothed data. The first column gives the scaling factor for individual seasons based on a total least square fit (tls), the 2nd column provides the 5-95% range of the scaling factors (note that ranges not encompassing zero show that the response to forcing is detectable, shown bold). Scaling factors that do not encompass 1 yield results that need to be significantly scaled to match the observations, indicated by an asterisk. Results using ordinary least squares (ols) estimates rather than total least squares (tls) estimates are shown in the third column (best guess for* β *and ratio of it to noise standard deviations).*

	Best guess scaling tls	5-95% range	Best ols and std dev
winter	**0.59**	[0.34, 1.05]	**0.39**, 3.12s
spring	**0.53**	[0.27 1.15]	**0.37**, 2.65s
summer	0.47	[-0.2 1.00]	0.19, 1.18s
fall	**0.23***	[0.00 0.71]	0.19, 1.36s
annual	**0.46***	[0.19 0.84]	**0.38**, 2.72s

Table SI3 lists the detection and attribution results. Results with scaling significantly less than 1 (as in fall and annual data) suggest that the reconstructions show signals of significantly smaller magnitude than the multimodel mean. This may be either due to the models having too strong forcing, a stronger response than observed, or due to the reconstruction, which may not resolve the full amplitude of the response (see discussion of reconstruction).We have

DOI: 10.1038/NGEO1057

assessed the robustness of our attribution results by using the residual variability in the reconstruction rather than the intermodel variability. Results show that the response to volcanism would be detectable based on the residual variability, but not on climate model variability (note that volcanism was robustly detectable on hemispheric mean scales[7]). However, the shorter record of the residual makes attribution results for anthropogenic and solar signals affected by very low estimates of independent degrees of freedom in the record when using fingerprints which vary slowly in time such as solar or greenhouse gases (details see above.).

Results show that for annual mean data, a best estimate of 71% (5-95% uncertainty 29-131%) of the warming since the little ice age was forced; for winter data the best estimate is 75% (44-133%) and for spring, 53% (27-116%).

Detection results generally are less robust and weaker when records are shorter, such as preindustrial only, and only from 1675 onwards:

Results 1675 -1996: The response to all forcings combined is detectable in winter (ordinary least squares, OLS, and total least squares, TLS), but not in annual mean data. Summer data show no detectable response to all forcings combined, nor separable responses to individual forcings. Thus the response to solar forcing is not detectable.

This slight deterioration of results with the shorter period illustrates that results using multiple regression need a reasonably long analysis period to be able to separate the response from different external forcings from each other and from internal variability on a relatively small scale such as Europe. Overall, results based on the shorter period are consistent with findings based on the entire reconstruction.

Results using data only prior to 1900: The response to all forcings combined prior to the 20th century is detectable only in summer, (OLS and TLS, both from 1500 and 1675 onwards). Individual forcings cannot be separated based on EBM results and model fingerprints available, although, again, there is a borderline solar signal in the 1500-1900 period emerging in the multi regression.

Although the better detectability and stronger correlations between models and reconstruction in preindustrial summer suggest that aerosol uncertainty in the models might make detection more difficult detection in that season during the 20th century (with aerosol forcing playing a role in suppressing warming in some models, while not even incorporated in ECHO-G, see above), analysis of the timeseries using only models with aerosol forcing did not improve the agreement between models or with the observations.

Interannual response to volcanic forcing

Volcanic forcing:
In order to determine to what extent robust interannual responses to volcanism are observed, we use recent reconstructions of volcanic forcing to define the timing of volcanic eruptions[21] and to determine the strongest eruptions, which are then used in a superposed epoch analysis. The reconstruction of volcanic forcing is based on sulphate peaks in a number of ice cores, both from Antarctica and Greenland, which are calibrated to represent aerosol optical depth (AOD) in a variety of latitude bands. We use the NH latitude band (30-90N) from one reconstruction[21] for our analysis. The analysis involves the following steps:

Detection of volcanic forcing: We used an analytic method to detect the timing of volcanic eruptions in the time series in order to be able to perform multiple sensitivity tests. For each season or annual mean, the difference between one year's or season's aerosol forcing and the next year's/season's forcing has been calculated. Where a strong

increase occurs between year n and n+1, year n is considered the time of a volcanic eruption. The threshold for the size of the increase in volcanic aerosols optical depth is set so that about 15 to17 eruptions are found between the period 1675 and the present (note that this means that the threshold varies slightly seasonally, part of which may be due to the default for the eruption data for unknown eruptions being set to January[21]). The eruptions that are found by this method largely coincide with known eruptions and are largely (but not completely) robust between seasons (see Figure 1).

We have also used in comparison the group of 15 large tropical eruptions analyzed in ref. 22, and a hand-selected number of volcanic eruptions with possible effect on Europe based on reference 21. All three records are compared in table SI4.

DOI: 10.1038/NGEO1057

Table SI4:Year of strong volcanic eruptions detected with three different methods: both left columns: automatic detection based on Crowley's[21] AOD index for summer (leftmost) and winter (2nd column; note that year given corresponds to year prior to an AOD peak, not necessarily eruption year). The value in brackets gives the known date of eruption compared to estimated pre-eruption year, and for known strong eruptions, the location; x indicates an extra tropical event. Third column: year preceding strong NH volcanic events suggested by Crowley (period of estimated peak stratospheric perturbation in mid-to high latitude given in brackets); a selection of strong tropical eruptions that should influence the subsequent winter substantially are shown in bold; right column: dates of 15 strong tropical eruptions, based on suggestion by Ammann and Naveau[16] and used in ref. 22.

Automatic estimate Summer	Automatic estimate winter	Crowley manual (1600 onwards)	Ammann manual
	1508		
1514 (1515)			
1525 (1526)	1525		
	1536		
1553			
1560 (1561)	1560,1561		
	1572,1573		
1584 (1585) x	1584,1585	1585 (8/1585)	1586
			1595
1599 (Quelcaya, 1600)	1599,1600	1600 (2/1601)	1600
1613 x			
1640 x; 2nd eruption 1641, Philipines)	1640	1640 (7/1641)	1641
1644 (1645)			
1661 (1662)			
1667 x	1667	1666 (10/1667)	
1672 (Indonesia, 1673)	1673	**1673** (1/1674)	1673
	1680		
1693? (1694, 2nd 1696)	1694	1695 (6/1696)	
1730 (1731-Lanzarote) x			
1739 (Japan) x	1739		
1783 (Laki) xx	1783		
1795 (1796) x			
1808 (1809)	1809	**1809** (2/1810)	1809
1814 (Tambora, 1815)	1815	**1815** (1/1816)	1815
			1822
1831	1831	1832 (6/1832)	1831
1834 (Cosiguina)	1835	1834 (8/1835)	1835
	1872		
1883 (Krakatau)	1883	**1883** (5/1884)	1883
1902 (Caribbean)	1902	**1902** (4/1903)	1902
1906 (Kamchatka, 1907) x	1907	1906 (5/1907)	
1911 (Caribbean 1912) x	1912	1911 (7/1912)	
1923 (unclear origin x?)	1924		
			1963
1981 (1982 El Chichon)	1982	**1982** (3/1983)	1982
1991 Pinatubo	1991	**1991** (2/1992)	1991

Superposed EPOCH analysis: The response to volcanic eruptions has been analyzed in a superposed epoch analysis (see methods section), which is described in the body of the paper. The analysis can be performed for both models and reconstruction. Figure SI4 expands on figure 3 in the paper, showing both years following the eruption. As spatial averaging reduces climate variability, the most powerful tests for significance of the climate impact are tests based either on the unusualness of the spatial mean change relatively to randomly picked years, or based on the similarity of the patterns between model and observations.

The main results of all these studies are given in the body of the paper, and details and sensitivity tests are listed in table SI5 below. If more eruptions are used (i.e. a lower threshold for aerosol increase in step 1 above), minor eruptions start being included in the analysis, leading to weaker results of the epoch analysis. On the other hand, if the threshold for including an eruption is set too high, then the number of years is not large enough to allow separation of the signal from noise, particularly for the highly variable winter season. This indicates that although volcanic eruptions cause significant impacts on European climate, their effect competes with substantial climate variability, which is why it is not always readily identifiable and requires a large number of eruptions to emerge. If using 15 selected tropical eruptions from ref 22 (which yields 10 post 1675, see also ref. 16), the winter response can only be marginally detected in the first year following the eruption. However, if 6 strong tropical eruptions that are timed such that peak stratospheric forcing would be expected for winter (table SI3) are used, their response can be detected more significantly in year one following the eruption (see figure SI5; note the difference in amplitude between models and reconstruction in that case). Significant winter warming occurs in most cases in either the first or the second year following peak forcing. When the automatically detected spikes (Table SI4) were separated into extra tropical and tropical events, resulting winter patterns again become noisy and change dependent on excluding/including individual events. All this demonstrates that while volcanic eruptions tend to cause winter warming, this response competes with very significant winter variability and is not robust based on a small number of eruptions.
In contrast to the noisy winter results, the fingerprint of summer cooling is detected irrespective of details of the analysis (Table SI5). Results for spring and autumn are not significant, with patterns showing a weakened winter warming in spring, and weak cooling in fall.

DOI: 10.1038/NGEO1057

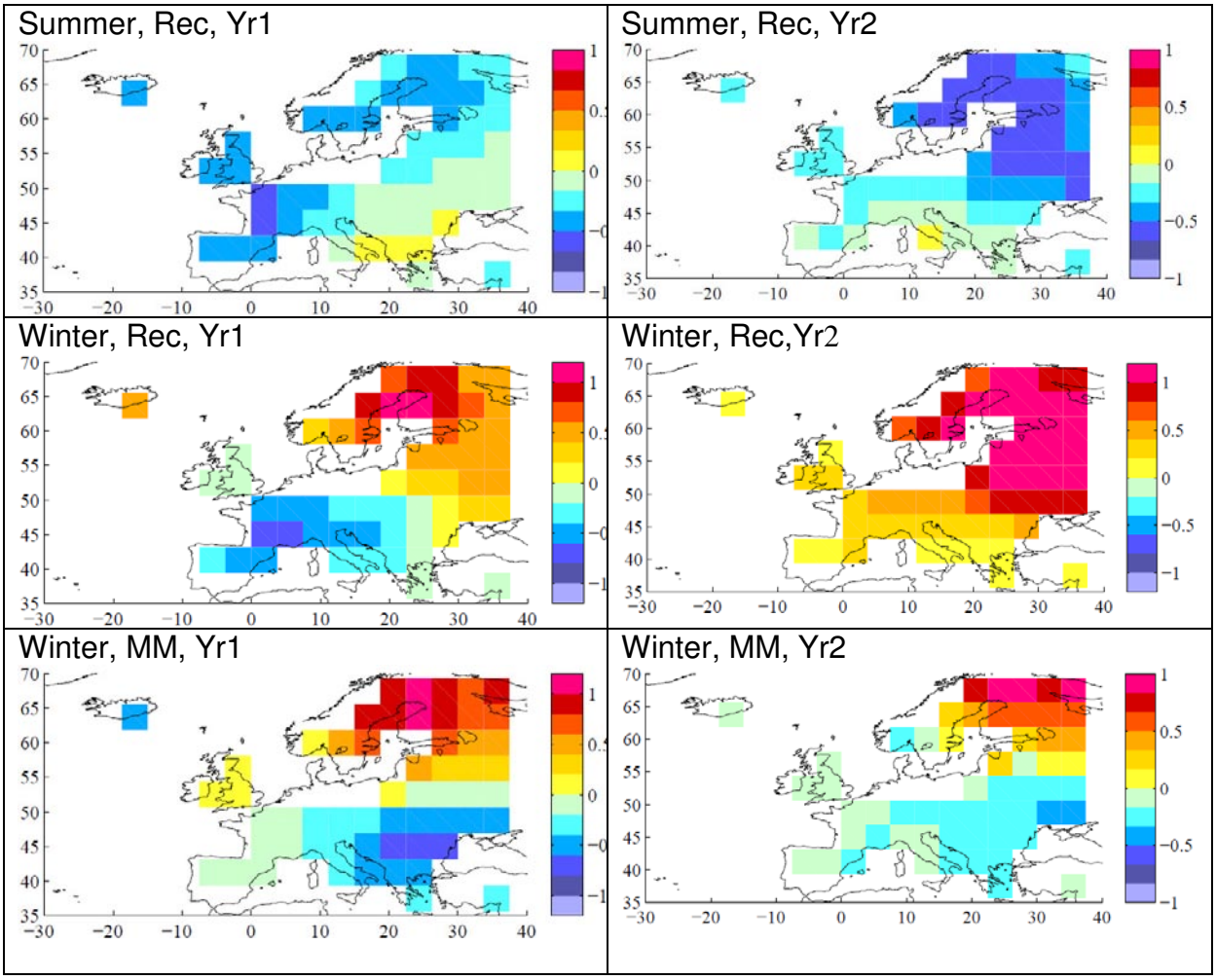

Figure SI4: *Temperature pattern following volcanic eruptions. Seasonal average temperature anomaly [K] following volcanic forcing in reconstruction year 1 and 2 following the estimated year before the eruption (labelled as year 0). Top panel shows summer, middle panel winter (both from reconstructions 'Rec', analyzing the period past 1675). The multimodel mean (MM) for winter is shown in the bottom panel. The EPOCH analysis is based on 17 volcanic eruptions with the largest AOD spikes in the NH from 1675-2000.*

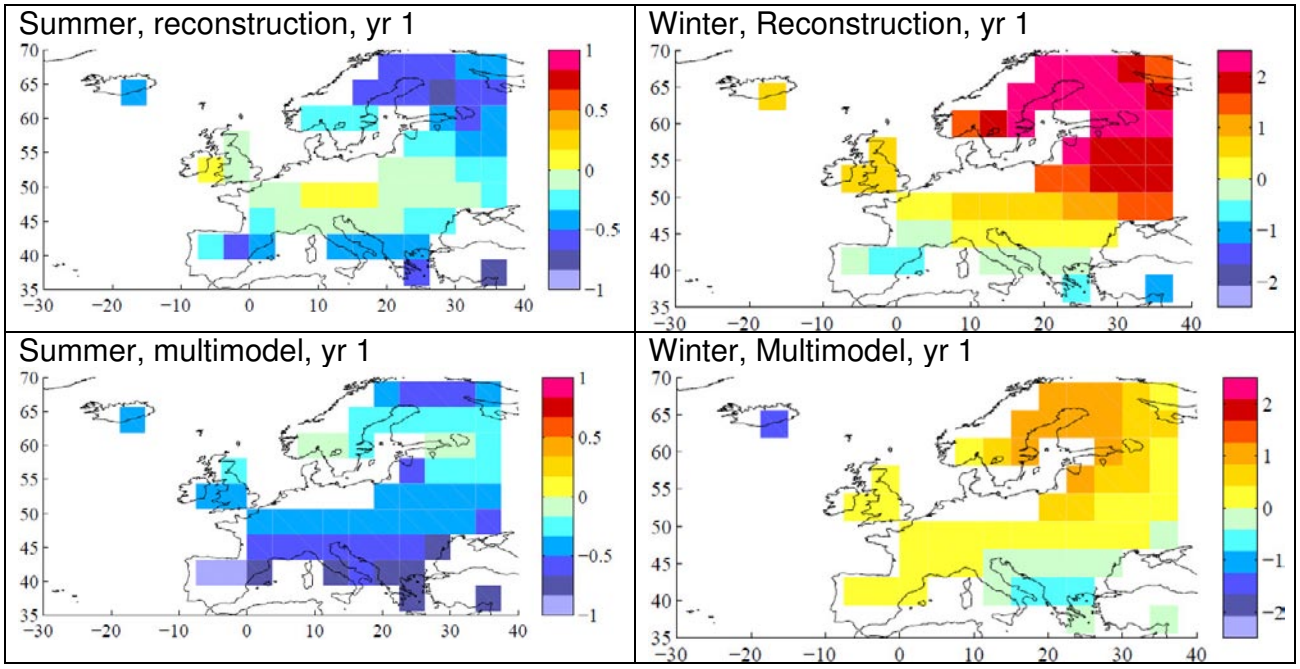

***Figure SI5**: As Figure 3 of main text, but using instead 6 tropical volcanic eruptions with peak forcing in winter, table SI4. Results are similar if using the eruptions used in ref. 23, either from 1675 on or over the entire 500 yrs (not shown). Note that the colour scale in winter is different from that in SI3.*

Table SI5: *Results from superposed epoch detection analysis for volcanic signals. The leftmost column lists the number of eruptions since 1675, with the number over the entire period (1500-1999) given in brackets. The 2nd column gives results for a significance test of area averaged EPOCH patterns from the reconstruction (results are listed if mean is significantly stronger than 95% of area averages picked by random eruptions) along with the sign of the change. Column 3 lists the signal to noise ratio of the regression of the observed epoch result onto the model epoch fingerprint, with uncertainty ranges determined by regression of random years from the reconstruction onto the multi model fingerprint. Results that are significant at the 5% level (one-sided) are printed bold.*

	Number of eruptions (full)	10% Significant in area mean	Signal/noise ratio for 2-yr fingerprint
wi	17 (33)	Yr 2 warming	**2.2 < 5%**
	21 (37)	Yr 2 warming	1.76
	10(15) Ammann	Yr 2 warming	Small, yr1 sign
	6 Crowley	Yr 1,2 warming	Small, yr1 sign
sp	17 (43)	no	small
	10(15) Ammann	no	small
Su	17 (31)	Yr1, Yr2 cooling	**3.3 << 5%**
	10(15) Ammann	Yr1, Yr2 cooling	**2.9 << 5%**
	6 Crowley	No	**2.4 <5%**
fall	15 (28)	No	small
	10(15) Ammann	No	small

DOI: 10.1038/NGEO1057

Response to periods of intense volcanism:

Periods of heavy volcanism have been chosen as periods with a 15yr average aerosol loading at peak values, starting from year 1 of the 15 yr period to capture interdecadal variations associated with volcanism. The highest values that are non-overlapping have been selected. We expect that these periods should be colder due to heavy volcanic forcing since the ocean retains memory of volcanic eruptions. We find that winters of peak volcanic forcing, which occurs over the 4 periods starting in years 1590, 1630, 1804 and 1830 corresponds to unusually cold winter conditions (see Figure SI7 for pattern; note that patterns are shown as absolute values rather than relative to prior years due to small sample). However, as this analysis does not allow for a separation of the role of different forcings, it does not prove that the Little Ice Age cold winters were caused by volcanism.

Response to solar forcing

Based on the multiregression result for European seasonal temperature described in the body of the paper, there is some evidence for solar forcing being detectable in summer, but the result was sensitive to the analysis period. However, our knowledge on forcing and response is not equally robust between forcings. For example, greenhouse gas forcing is far better constrained than solar forcing, and volcanic forcing could be argued to be more robust than solar as well (e.g. ref. 23). To further investigate to what extent our results are robust to first identifying the more robust forcing responses, we used a stepwise regression[24], which first estimates the response to better constrained forcings, such as greenhouse gases and volcanism, and estimates solar forcing from the residual. To enhance power, we also use the spatial pattern. This is equivalent[24] to using a solar forcing fingerprint in time that is orthogonalized to that of anthropogenic forcing (see figure SI6; note that this essentially removes the anthropogenic component from the timeseries and resulting spatial pattern prior to a regression on the solar timeseries). This approach is consistent with the greater confidence in the shape and size of the anthropogenic forcing[24]. Note that if data up to 1950 are used, the only difference between the orthogonal and original regressor is that the long-term trend is no longer visible (Figure SI6). The resulting regression pattern was compared with patterns obtained by regression of a random time series of the same autocorrelation as the solar response time series onto the reconstruction data (see figure SI6 for some examples). If the summer reconstruction over the period 1500 to 1950 is used, then the response pattern to solar forcing is significantly warmer in the area average than that obtained from random time series in the reconstruction. If the same is done to model data, then the result is warming, which is not significant, but its pattern (Figure SI7) projects more strongly on that from the reconstruction than 90% of the cases where a random timeseries was regressed on both models and reconstruction. Similar results are obtained if the timeseries is analyzed until 1996 rather than 1950. In contrast, using the shorter period back to 1675 provides insignificant results.

However, if the solar pattern is orthogonalized to both the anthropogenic and volcanic EBM fingerprints (effectively removing the contribution by both forcings to the reconstruction prior to analysis) the response pattern is no longer detectable. This raises concerns that despite a low correlation between the EBM response to solar forcing and volcanism (0.11), some degeneracy may remain between both, particularly given that that both solar forcing and volcanism tend to cool the period termed the Little Ice Age. Thus, the detection of a solar response in European summer temperatures remains uncertain.

Regressions of the solar forcing timeseries on European temperatures in other seasons than summer (JJA) show no evidence for a detectable solar signal, and larger samples and model

simulations with individual forcingsare needed to assess the possibility of a dynamical response to solar forcing in the cold season as discussed in the literature[25].

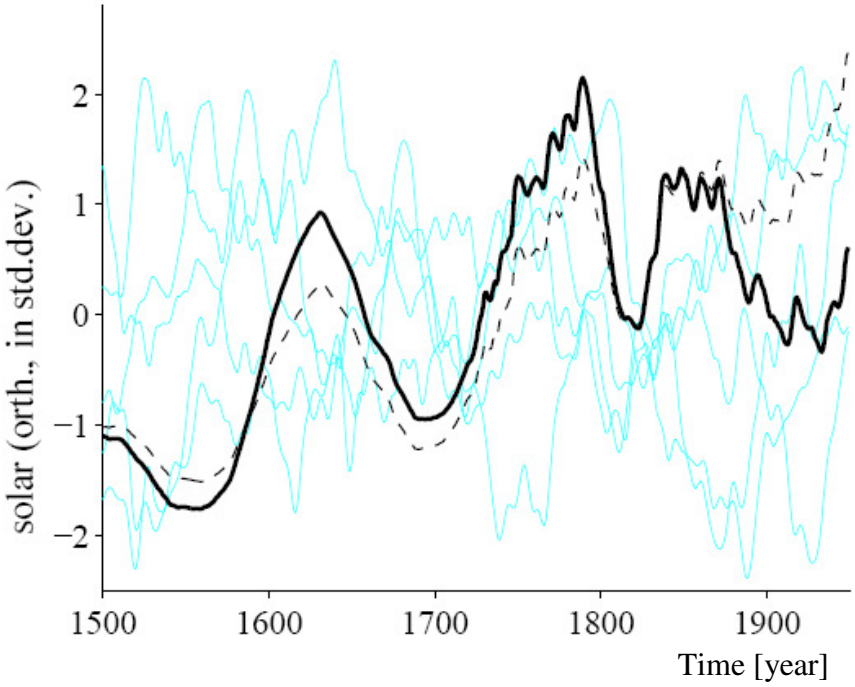

Figure SI6: *EBM response of NH mean temperature to solar forcing raw (dashed) and orthogonalized to anthropogenic forcing (solid). Both are scaled to a standard deviation of unity. In blue, realizations of a random time series of the same autocorrelation as solar forcing are shown, which are used for significance testing.*

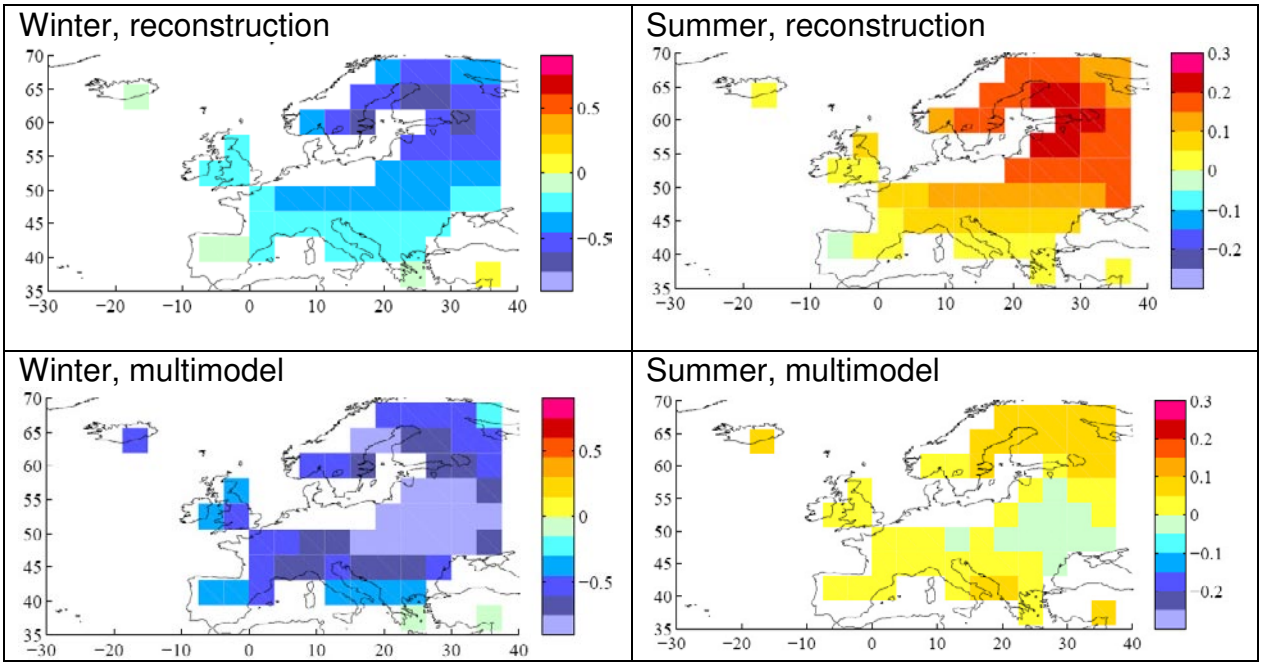

***Figure SI7**: Winters with strong volcanic forcing from multiple eruptions, in models (left bottom) and reconstruction (top, left) and regression of solar forcing on summer temperature in reconstructions (top, right) and multi model mean (bottom right).*

DOI: 10.1038/NGEO1057

References

1. Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M., & Wanner, H. European seasonal and annual temperature variability, trends and extremes since 1500. *Science* **303**, 1499-1503 (2004).
2. Luterbacher, J., et al. The Exceptional European warmth of Autumn 2006 and Winter 2007: Historical context, the underlying dynamics and its phenological impacts. *Geophys. Res. Lett*. **34**, L12704 (2007)
3. Xoplaki, E., et al. European spring and autumn temperature variability and change of extremes over the last half millennium. *Geophys. Res. Lett*., **32**, L15713, doi:10.1029/2005GL023424 (2005).
4. Mitchell, T., & Jones, P. D. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol*., **25**, 693–712 (2005).
5. Brohan P., Kennedy, J.J., Harris, I., Tett, S.F.B. & Jones, P.D. Uncertainty estimates in regional and global observed temperature changes: a new data set from 1850. *J Geophys. Res. Atmos*. **111**: D12106 (2006).
6. Küttel, M., et al. Testing a European winter surface temperature reconstruction in a surrogate climate, *Geophys. Res. Lett.* **34**, L07710 (2007).
7. Hegerl G. C. et al. Detection of human influence on a new 1500yr climate reconstruction. *J. Climate* **20**, 650-666 (2007).
8. Min, S., Legutke, S., Hense A., & Kwon, W. Internal variability in a 1000-yr control simulation with the coupled climate model ECHO-G -I. Near surface temperature, precipitation and sea level pressure. *Tellus*, **57A**, 605-621 (2005).
9. von Storch, H., et al. Reconstructing past climate from noisy data. *Science* **306,** 679–682 (2004).
10. Crowley, T. J. Causes of climate change over the past 1000 years, *Science*, **289**, 270-277 (2000).
11. Zorita, E., et al. Climate evolution in the last five centuries simulated by an atmosphere-ocean model: global temperatures, the North Atlantic Oscillation and the Late Maunder Minimum. *Meteorol. Zeitschrift*, **13**, 271-289 (2004).
12. González-Rouco J., Beltrami H., Zorita E., & Stevens, M. Borehole climatology: a discussion based on contributions from climate modeling. *Clim. Past,* **5**, 97–127 (2009).
13. Tett, S. F. B., et al. The impact of natural and anthropogenic forcings on climate and hydrology since 1550. *Clim. Dyn*. **28**, 3–34 (2007).
14. Johns, T. et al., Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios. *Clim. Dyn.* **20**, 583–612 (2003).
15. Ammann, C. M., Joos, F., Schimel, D., 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. Nat. Acad. Sci.* **104**, 3713-3718 (2007).
16. Ammann C.M. & Naveau, P. Statistical analysis of tropical explosive volcanism occurrences over the last 6 centuries, *Geophys. Res. Lett*. **30**, 1210, doi:10.1029/2002GL016388 (2003).
17. Lean, J., Wang, Y. & Sheeley, N. 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. Let*. **29**, 2224 (2002).
18. Foukal, P., North, G. & Wigley, T. M. L. A stellar view on solar variations and climate. *Science* **306**, 68-69 (2004).
19. Solanki, S. K., & Krivova, N. A. Solar irradiance variations: from current measurements to long-term estimates. *Solar Phys.* **224**, 197-208 (2004).

20. Allen, M. R. & Stott, P. A. Estimating signal amplitudes in optimal fingerprinting, Part I: Theory. *Clim. Dyn.* **21**, 477–491 (2003)
21. Crowley, T. J. et al. Volcanism and the Little Ice Age. *PAGES News* **16**, 22-23 (2008)
22. Fischer, E.M., et al. European climate response to tropical volcanic eruptions over the last half millennium, *Geophys. Res. Lett*. **34**, L05707 (2007).
23. Gray, J.L. et al. Solar Influences on Climate. *Rev. in Geophys*, 48, RG4001, doi:10.1029/2009RG000282. (2010)
24. Hegerl, G. C. & Allen, M. R. Origins of model-data discrepancies in optimal fingerprinting. *J. Climate* **15**, 1348-1356 (2002)
25. Shindell, D., G. Schmidt, R. Miller, and M. Mann (2003), Volcanic and solar forcing of climate change during the preindustrial era. *J. Climate* **16**, 4094–4107.