LETTER

doi:10.1038/nature11969

Recent temperature extremes at high northern latitudes unprecedented in the past 600 years

Martin P. Tingley[1] & Peter Huybers[1]

Recently observed extreme temperatures at high northern latitudes[1–7] are rare by definition, making the longer time span afforded by climate proxies important for assessing how the frequency of such extremes may be changing. Previous reconstructions of past temperature variability have demonstrated that recent warmth is anomalous relative to preceding centuries[2,8,9] or millennia[10], but extreme events can be more thoroughly evaluated using a spatially resolved approach that provides an ensemble of possible temperature histories[11,12]. Here, using a hierarchical Bayesian analysis[13,14] of instrumental, tree-ring, ice-core and lake-sediment records, we show that the magnitude and frequency of recent warm temperature extremes at high northern latitudes are unprecedented in the past 600 years. The summers of 2005, 2007, 2010 and 2011 were warmer than those of all prior years back to 1400 (probability $P > 0.95$), in terms of the spatial average. The summer of 2010 was the warmest in the previous 600 years in western Russia ($P > 0.99$) and probably the warmest in western Greenland and the Canadian Arctic as well ($P > 0.90$). These and other recent extremes greatly exceed those expected from a stationary climate, but can be understood as resulting from constant space–time variability about an increased mean temperature.

Exceptionally high temperatures have recently been observed in Russia[1–4], Greenland[5,6] and other locations at high northern latitudes[7]. Palaeoclimate reconstructions demonstrate that these temperatures are anomalous[2,8–10], but ascertaining whether they are unprecedented requires a treatment beyond that supported by standard measures of uncertainty[11,12]. At issue is how warm it could have been at each location and each year in the past, given the uncertainty in each temperature estimate. For example, a temperature $T_0 = 33\,^{\circ}\mathrm{C}$ is almost certainly higher than the ith realization of $T_i = 30 + \varepsilon_i$, where ε is a random variable drawn from a normal distribution with zero mean and unit variance, but it is unlikely to be higher than the maximum of 1,000 independent realizations of T_i.

The probability that a given interval contains the most extreme temperature in the span of a reconstruction is generally quantified using ensemble-based reconstruction methods because they can be used to estimate simultaneous, or pathwise, uncertainty intervals[11,15], and provide direct probabilistic assessments of extremes[12]. Ensemble-based reconstructions have recently been used to evaluate extremes in spatially averaged temperature using bootstrap methods[16], Bayesian principal component regression[17] and realizations drawn from global climate model simulations[18,19].

Here we use a hierarchical Bayesian model[13,14,20–22] to perform a $5^{\circ} \times 5^{\circ}$ spatial resolution analysis of the past 600 years of summer temperatures at high northern latitudes (see Methods). This approach gives a spatially resolved ensemble of temperature reconstructions, each of which is equally likely given the available instrumental and proxy data (see Methods). This approach also allows estimation of temperature and its uncertainty at locations without instrumental or proxy observations. We can therefore assign probabilities to years being warmest or coldest at each spatial location[11,12] and assess recent extremes in the context of space–time climate variability.

The reconstruction is based on April–September averages of the Climate Research Unit's gridded instrumental data set[23] and 125 annually resolved proxy time series: 18 ice-core $\delta^{18}O$ series, 11 log-transformed annual varve thickness series from lake-floor sediment cores and 96 gridded maximum latewood tree-ring density series. We infer April–September temperature anomalies from 1400 to 2011 at all grid boxes between 45° N and 85° N that contain some fraction of land. The requirement that the proxy series be annually resolved, coupled with the need for sufficient coverage to produce a spatially complete reconstruction, necessitates that we limit the study to the past 600 years.

For temperature anomalies averaged over land areas north of 45° N, the years 2005, 2007, 2010 and 2011 are each warmer than all other years in the 1400–2004 interval ($P > 0.95$; Fig. 1a), whereas on decadal timescales, intervals subsequent to 1996 are warmer than all previous intervals ($P > 0.95$; Fig. 1b). These results support and extend previous findings regarding unprecedented warmth[2,9,10] by explicitly considering whether a given anomaly is higher than those for all other years. Spatially averaged temperature estimates become more uncertain at earlier times, particularly before 1850 when the longest instrumental records drop out of the analysis (Fig. 1a and b). The unweighted average across the available instrumental observations closely agrees with the spatially averaged estimates from all data types back to about 1900, but not before. As the early instrumental observations are sparsely and heterogeneously distributed in space, the spatially complete reconstruction afforded by the model (see Methods) and the proxy indicators of temperature become increasingly important towards earlier times.

Centennial-timescale variations can be characterized by linear trends (Fig. 1c). The linear rate of warming in the spatial-mean time series over the interval 1912–2011 is 1.06 °C per century (90% uncertainty is 1.03 °C to 1.08 °C). With $P > 0.99$, the largest 100-year warming trend is centred after 1900, indicating that recent rates of warming are unprecedented in the past 600 years. The 100-year cooling trend with the largest magnitude occurred between 1547 and 1646 at a rate of −0.71 °C per century (90% uncertainty is −0.88 °C to −0.54 °C). Interestingly, the warming trend has only recently exceeded the magnitude of this earlier cooling trend, such that if the reconstruction is curtailed at 2001, the cooling has a larger magnitude rate of change in 49% of the ensemble members. This early cooling trend, whose magnitude rivals that of the recent warming, also indicates that the proxy record is capable of capturing large, persistent temperature variations.

To investigate the distribution of extremes in both space and time, we calculate the probability that each year was warmest or coldest for each spatial location (Fig. 2). Whereas the 2000s had a high probability of containing the warmest year in many locations throughout northeastern Canada, Greenland, eastern Russian and the land areas surrounding the Bering Sea, the 1990s contained the warmest year predominantly in northwestern North America, and, to a lesser extent, central Russia (Fig. 2a). Averaging across the ensemble, we find that the warmest year occurs in 1990 or later for 60% of all locations, and in 2000 or later for 45%, demonstrating that the recent extreme warmth is widespread (Fig. 2b). There is no such clear trend in the occurrences of

[1]Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, Massachusetts 02138, USA.

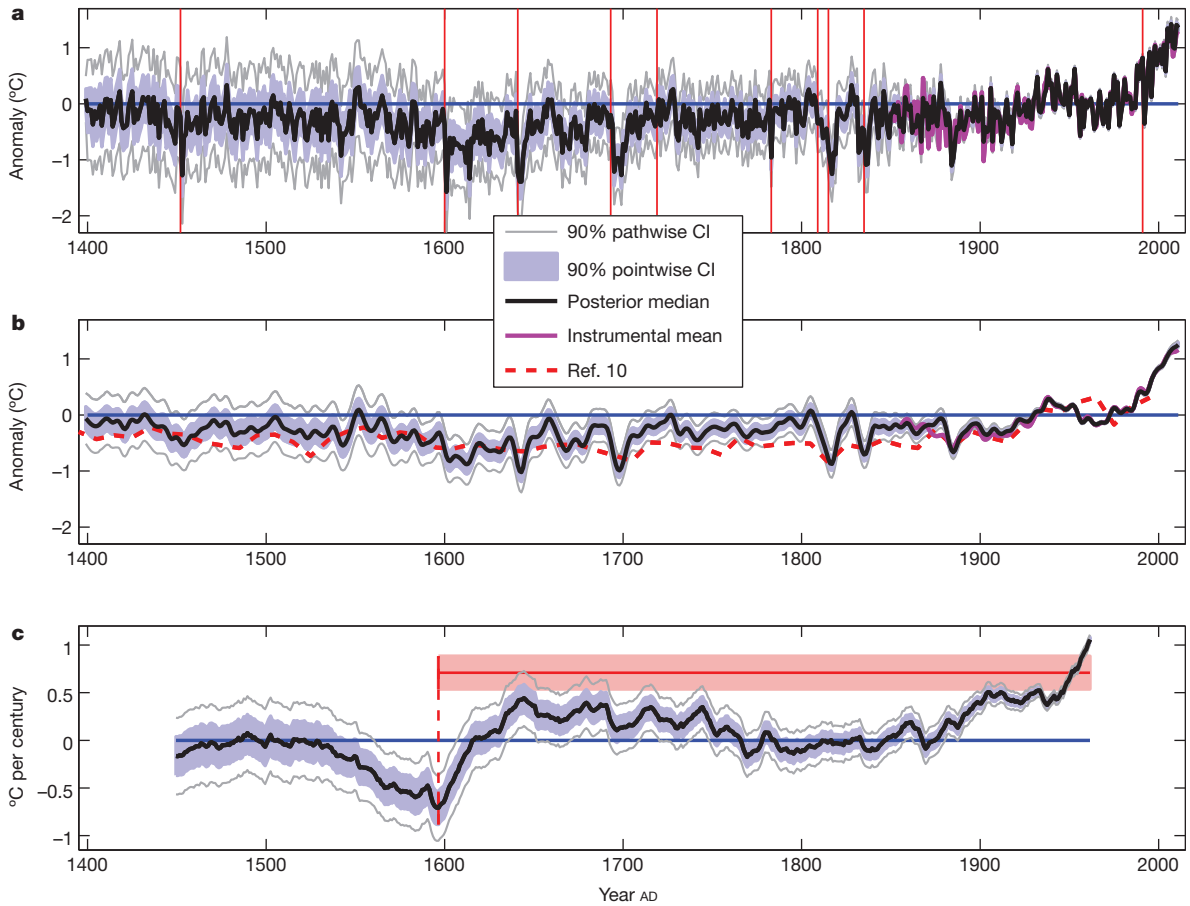

Figure 1 | Time series of temperature anomalies and centennial slopes. **a**, Average land temperature between 45° N and 85° N (black), 90% pointwise (blue shading) and pathwise (grey) credible intervals[20] (see Methods); the unweighted average of all available instrumental observations (magenta); and the ten largest volcanic eruptions in the 1400–2011 interval according to ice-core sulphate concentrations[30] (vertical red). **b**, As in **a**, but individual ensemble members are first smoothed with a nine-year Hanning window, along with a separate reconstruction[10] (dashed red). **c**, As in **a**, but for linear trends calculated for overlapping 100-year intervals. To facilitate comparison between the recent rate of warming and earlier rates of cooling, the median and 90% pointwise uncertainty for the cooling centred at 1596 is also inverted in sign and extended across to the modern period (red lines and shading).

minima, which are clustered in times following volcanic eruptions, as well as during the cool 1600s. As observed elsewhere[24], incidences of record-breaking temperature maxima have been increasing in frequency over the past few decades, whereas the rate of minima in the summer average temperature has fallen close to zero.

At a more regional level, the summer of 2010 is very likely ($P > 0.90$) to have been the warmest in seven of the eight grid boxes in the region affected by the Russian heat wave (50° N–60° N, 35° E–55° E (ref. 1); Fig. 2a). Averaging temperatures over this region reduces the variability between ensemble members, and we then find that 2010 is the warmest summer in western Russia in more than 99.9% of the ensemble members. In contrast, the summer of 2003 is probably ($P > 0.66$) the warmest for only two grid boxes in western Europe (Fig. 2a), in agreement with other findings[2,9] that the 2010 Russian heat wave exhibited a broader spatial signature of extremes. We also find that 30 of the 47 continental grid boxes in western Greenland and the Canadian Arctic archipelago are more likely than not to have experienced their warmest conditions in 2010, in agreement with other more localized indications of temperature, runoff and surface mass balance in this region[5–7]. For the average over this region, 2010 has the warmest conditions in 93% of the ensemble members.

The robustness of these results is indicated by the fact that using subsets of the proxy data gives similar structure with respect to the occurrence of extreme events. For instance, estimates based on only the ice-core and varve proxies, with model parameters derived using the complete data set (see Methods), give very similar results to inferences from all data sources, which are dominated by tree-ring densities before 1850 and by instrumental observations thereafter (Fig. 2c). Both reconstructions show substantially elevated probabilities of maxima occurring in recent decades, very low probability of maxima in 1600–1850 and slightly elevated probabilities in the first two centuries of the reconstruction (Fig. 2b and c). Estimates using only tree-ring-density records result in a distribution of extreme years that is essentially unchanged, provided the comparison extends only to 1960, after which these proxies systematically underestimate temperature anomalies—as expected, given the tree-ring divergence phenomenon[25]. More generally, we find that the various temperature indicators used in this study agree with each other given their respective uncertainties, excepting the tree-ring-density records after 1960 (see Methods and Supplementary Information).

The increased frequency of warm extremes in recent decades (Fig. 2b and c) may result from increased average temperature, increased variance or changes in higher-order moments of the temperature distribution[26]. To explore the underlying cause of the recent increase in warm extremes, we use the Bayesian model and its data-derived parameters to simulate both temperature anomalies and observations of those anomalies having noise properties consistent with the instrumental data (see Methods). The distribution of simulated temperature anomalies over the past 20 years is biased substantially lower than estimates using the actual data (Fig. 3a). Shifting the mean of the simulated anomalies to match that observed over the past 20 years, which is 1.16 °C higher than the 600-year average, results in much closer agreement between the distributions of simulated and estimated temperatures (Fig. 3a), although the five highest individual

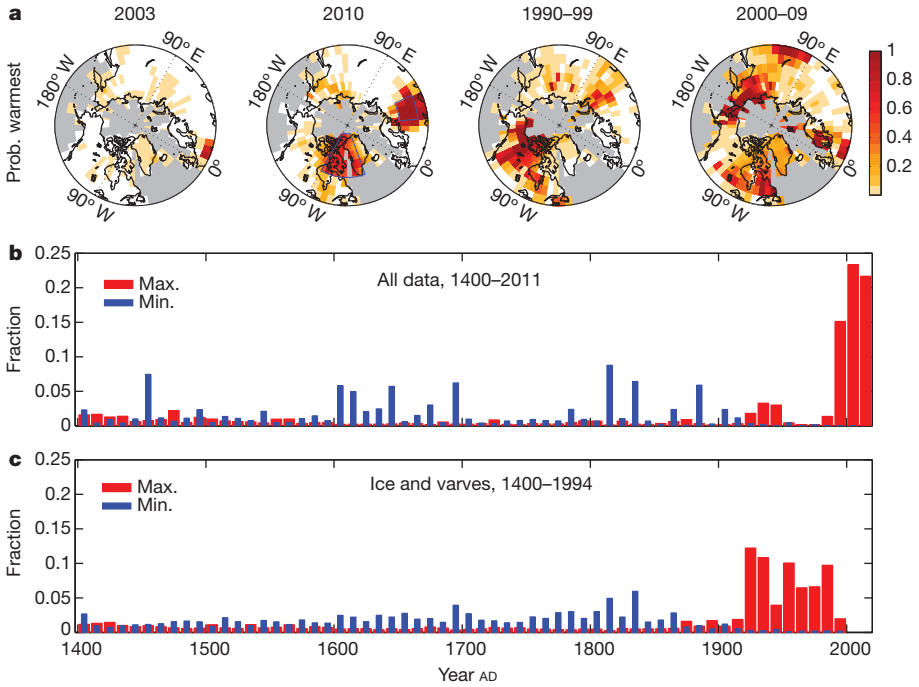

Figure 2 | Warm and cold extremes. **a**, The proportion of draws (see Methods) for which 2003 and 2010 were warmest, and for which the warmest year fell in the 1990s and 2000s. White shading indicates zero. **b**, The fraction of all locations for which years were warmest or coolest, averaged across all ensemble members. Results are binned by decade, except for the last interval, which contains only 2010–2011. **c**, As in **b**, but for a reconstruction that uses only the ice-core and lake-varve series and covers the interval 1400–1994.

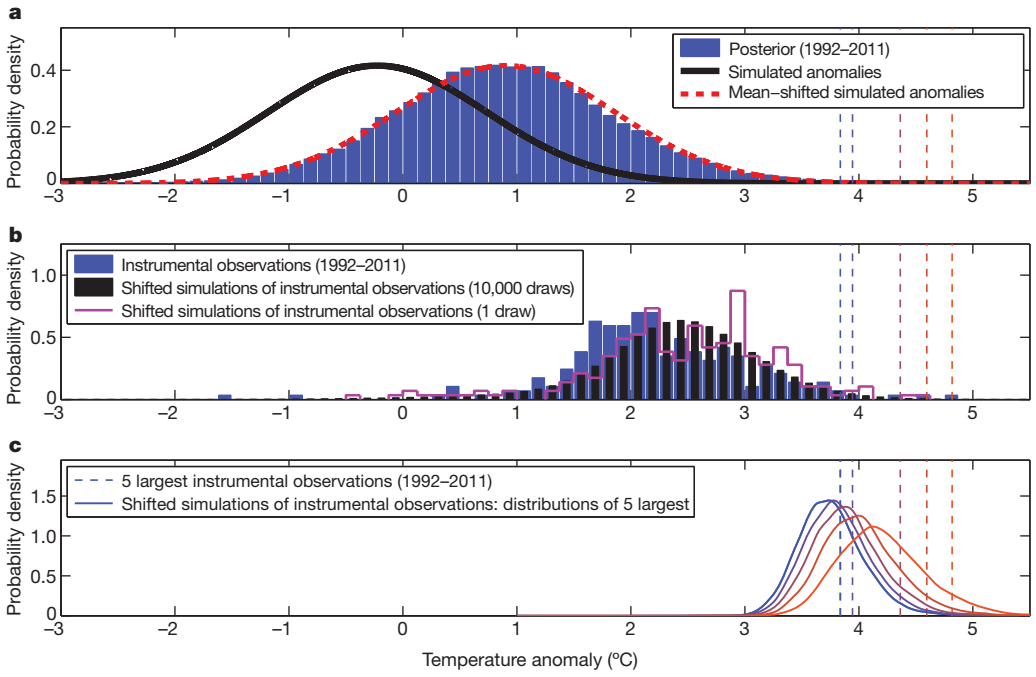

Figure 3 | Histograms of temperature anomalies and instrumental maxima for the period 1992–2011. **a**, Histogram of temperature anomalies across locations, ensemble members and years for the interval 1992–2011 (blue); the simulated distribution of temperature anomalies, using median parameter values fitted over 1400–2011 (black) and after shifting its mean to that of the 1992–2011 anomalies (dashed red). Vertical lines are repeated in each panel and correspond to the 3.8 °C anomaly near Moscow in 2010 and the four other even more extreme values in the 1992–2011 interval which, from lowest to highest, are from Siberia in 2007, Svalbard in 2006 and two locations in northern Canada in 1998. **b**, Histogram of the maximum value of the instrumental observations at each location over the 1992–2011 interval (blue), and the distribution of the maximum values according to 10,000 realizations (black) and a single (magenta) realization based on the simulations with the mean shifted as in **a**. **c**, Distributions of the five largest simulated instrumental observations across space and time in the 1992–2011 interval, with the mean shifted as in **a**.

instrumental anomalies during the past 20 years remain far in the tail of the simulated distribution.

Analogous to the need for pathwise uncertainties when evaluating whether particular years featured the warmest conditions (Fig. 1), we must also take into account that the five highest observed temperature extremes are selected from a range of locations. Shifting the mean and selecting the largest simulated instrumental observations in both time (Fig. 3b) and space (Fig. 3c) results in distributions consistent with the observed extremes, each of which falls below the 95th percentile of the corresponding distribution from the simulations (Fig. 3c). To first order, therefore, the highest temperature observations over the past 20 years can be accounted for by a shift in the average temperature and by considering the distribution of maxima across both space and time.

We find a similar relationship between the mean and extremes when performing the same analysis for earlier 20-year intervals between 1852 and 1991 (see Supplementary Information), indicating that the observed extremes require no change to the distribution of temperature anomalies other than to the mean, at least on the temporal and spatial scales of this analysis. Applying the same analysis to 100-year intervals of each proxy data set gives similar results, although with more scatter than the results shown in Fig. 3. There are two exceptions where the distributions of positive extremes do not follow the mean: over the most recent 100-year interval of tree-ring-density observations, positive extremes are biased low on account of the divergence phenomenon, and there is a general tendency for the largest log-transformed varve observations to be biased high relative to the mean-shifted simulations, suggesting that the log transform is an insufficient scaling for these largest values.

During the past 20 years, 61% of all high northern latitude locations are more likely than not to have experienced warm temperatures that are unprecedented in the past 600 years, and 25% of northern locations are very likely ($P > 0.95$) to have experienced high temperatures that are without precedent. Assessment of these recent temperature extremes in the context of space–time variability and observational error allows for a simple, cohesive explanation in terms of an increase in mean temperature. Specifically, the Bayesian model of variability in conjunction with a mean warming reproduces the general distribution of temperature anomalies (Fig. 3a), the maximum instrumental observation at each location (Fig. 3b) and the highest instrumental temperature observations over space and time (Fig. 3c). These results are consistent with analyses of global instrumental records demonstrating that recent incidences of extreme temperature are a result of an upward shift in the mean of the temperature distribution[24] but that the variance of that distribution has not significantly changed[27]. Further work is needed, however, to assess whether extremes also follow mean temperatures at smaller temporal and spatial scales, such as those associated with local weather.

METHODS SUMMARY

Observations are first re-expressed as anomalies with respect to the full time interval spanned by each data type[28] and the ice-core and log-transformed lake-varve series are additionally standardized to unit variance. The Bayesian model[13,14] represents the temperature anomalies as a first-order autoregressive process in time with exponentially decaying spatial covariance. Each proxy type is separately modelled as linearly related to temperature anomalies with mean-zero normal noise, and instrumental observations are modelled as temperature anomalies with mean-zero normal noise.

The spatial mean time series is computed for each ensemble member by weighting the grid boxes by land area. Taking the 50th percentile at each year gives the best estimate of temperature, taking the 5th and 95th percentiles produces 90% pointwise uncertainty envelopes, and expanding the pointwise uncertainty envelopes to contain 90% of ensemble members in their entirety produces pathwise uncertainties[15,17,29] (Fig. 1). Estimates of temporally smoothed quantities and centennial-scale trends are derived analogously after applying these operations to each ensemble member in turn (Fig. 1). Probability statements concerning extremes are derived by calculating the fraction of ensemble members that feature warmest or coldest conditions for a given year or interval (Fig. 2). To assess the robustness of results, we perform reconstructions using only subsets of the data (Fig. 2c). To explore recent extremes, we simulate the expected variability of instrument-like observations in the absence of data (Fig. 3), using model parameters that result from the full analysis.

The analysis technique gives results at a level comparable to or better than other climate field reconstruction techniques[22], even when the underlying assumptions are not fully met[14]. Validation metrics similarly suggest that the simple assumptions underlying the analysis are an adequate statistical description of the data, whereas experiments based on predicting withheld instrumental observations from the proxies indicate that the uncertainty intervals have accurate coverage properties.

Full Methods and any associated references are available in the online version of the paper.

Received 1 September 2012; accepted 29 January 2013.

1. Dole, R. *et al.* Was there a basis for anticipating the 2010 Russian heat wave? *Geophys. Res. Lett.* **38,** L06702 (2011).
2. Barriopedro, D., Fischer, E., Luterbacher, J., Trigo, R. & García-Herrera, R. The hot summer of 2010: redrawing the temperature record map of Europe. *Science* **332,** 220–224 (2011).
3. Rahmstorf, S. & Coumou, D. Increase of extreme events in a warming world. *Proc. Natl Acad. Sci. USA* **108,** 17905–17909 (2011).
4. Otto, F. E. L., Massey, N., Oldenborgh, G. J. V., Jones, R. G. & Allen, M. R. Reconciling two approaches to attribution of the 2010 Russian heat wave. *Geophys. Res. Lett.* **39,** L04702 (2012).
5. van As, D. *et al.* Large surface meltwater discharge from the Kangerlussuaq sector of the Greenland ice sheet during the record-warm year 2010 explained by detailed energy balance observations. *Cryosphere* **6,** 199–209 (2012).
6. Tedesco, M. *et al.* The role of albedo and accumulation in the 2010 melting record in Greenland. *Environ. Res. Lett.* **6,** 014005 (2011).
7. Richter-Menge, J. & Jeffries, M. The Arctic. *Bull. Am. Meteorol. Soc.* **92,** S143–S160 (2011).
8. Overpeck, J. *et al.* Arctic environmental change of the last four centuries. *Science* **278,** 1251–1256 (1997).
9. 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).
10. Kaufman, D. S. *et al.* Recent warming reverses long-term Arctic cooling. *Science* **325,** 1236–1239 (2009); correction **327,** 644 (2010).
11. National Research Council. *Surface Temperature Reconstructions for the Last 2000 Years* (National Academies Press, 2006).
12. Tingley, M. *et al.* Piecing together the past: statistical insights into paleoclimatic reconstructions. *Quat. Sci. Rev.* **35,** 1–22 (2012).
13. Tingley, M. & Huybers, P. A Bayesian algorithm for reconstructing climate anomalies in space and time. Part 1: development and applications to paleoclimate reconstruction problems. *J. Clim.* **23,** 2759–2781 (2010).
14. Tingley, M. & Huybers, P. 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 (2010).
15. Craigmile, P., Guttorp, P. & Percival, D. Trend assessment in a long memory dependence model using the discrete wavelet transform. *Environmetrics* **15,** 313–335 (2004).
16. Li, B., Nychka, D. & Ammann, C. The 'hockey stick' and the 1990s: a statistical perspective on reconstructing hemispheric temperatures. *Tellus A* **59,** 591–598 (2007).
17. McShane, B. & Wyner, A. A statistical analysis of multiple temperature proxies: are reconstructions of surface temperatures over the last 1000 years reliable? *Ann. Appl. Stat.* **5,** 5–44 (2011).
18. Christiansen, B. & Ljungqvist, F. Reconstruction of the extra-tropical NH mean temperature over the last millennium with a method that preserves low-frequency variability. *J. Clim.* **24,** 6013–6034 (2011).
19. Christiansen, B. & Ljungqvist, F. The extra-tropical northern hemisphere temperature in the last two millennia: reconstructions of low-frequency variability. *Clim. Past* **8,** 765–786 (2012).
20. Gelman, A., Carlin, J., Stern, H. & Rubin, D. *Bayesian Data Analysis* 2nd edn (Chapman and Hall/CRC, 2003).
21. Berliner, L., Wikle, C. & Cressie, N. Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. *J. Clim.* **13,** 3953–3968 (2000).
22. Werner, J., Luterbacher, J. & Smerdon, J. A pseudoproxy evaluation of Bayesian hierarchical modeling and canonical correlation analysis for climate field reconstructions over Europe. *J. Clim.* **26,** 3 851–867 (2013).
23. Brohan, P., Kennedy, J., Harris, I., Tett, S. & Jones, P. Uncertainty estimates in regional and global observed temperature changes: a new data set from 1850. *J. Geophys. Res.* **2,** 99–113 (2006).
24. Hansen, J., Sato, M. & Ruedy, R. Perception of climate change. *Proc. Natl Acad. Sci. USA* **109,** E2415–E2423 (2012).
25. Briffa, K. *et al.* Reduced sensitivity of recent tree-growth to temperature at high northern latitudes. *Nature* **391,** 678–682 (1998).
26. IPCC. in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (eds Field, C. *et al.*) 1–19 (A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change, Cambridge Univ. Press, 2012).
27. Rhines, A. & Huybers, P. Frequent summer temperature extremes reflect changes in the mean, not the variance. *Proc. Natl Acad. Sci. USA* **110,** E546 (2013).

28. Tingley, M. A Bayesian ANOVA scheme for calculating climate anomalies, with applications to the instrumental temperature record. *J. Clim.* **25,** 777–791 (2012).
29. Krivobokova, T., Kneib, T. & Claeskens, G. Simultaneous confidence bands for penalized spline estimators. *J. Am. Stat. Assoc.* **105,** 852–863 (2010).
30. Gao, C., Robock, A. & Ammann, C. Volcanic forcing of climate over the past 1500 years: an improved ice core-based index for climate models. *J. Geophys. Res. D* **113,** D23111 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements Data analysis was performed on the Odyssey cluster supported by the FAS Science Division Research Computing Group at Harvard University. Funding for this work was provided in part by NSF grant ATM-0902374. We thank E. Butler, P. Craigmile, N. Cressie, M. Haran, B. Li, E. Mannshardt, K. McKinnon, D. Nychka, B. Rajaratnam, A. Rhines, D. Schrag and A. Stine for comments and discussions.

Author Contributions M.P.T. performed the analysis. Both authors contributed to the design of the analysis, the interpretation of results, and preparation of the manuscript.

Author Information Data and code are available from NOAA Paleoclimatology at ftp://ftp.ncdc.noaa.gov/pub/data/paleo/contributions_by_author/tingley2013/. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.P.T. (tingley@fas.harvard.edu).

METHODS

Data. Instrumental observations are derived from the gridded $5^\circ \times 5^\circ$ CRUTEM3v data set compiled by the University of East Anglia's Climate Research Unit (ref. 23; data available at http://www.cru.uea.ac.uk/cru/data/temperature/). We use all time series having at least ten years of complete monthly April–September data, that are poleward of 45° N, and that contain a non-zero fraction of land according to a $0.25^\circ \times 0.25^\circ$ land–sea mask (ref. 31; data available at http://ldas.gsfc.nasa.gov/gldas/GLDASvegetation.php). To avoid introducing spurious structure in the time series of spatial standard deviations, which could result from the short 1961–90 reference period used in standardizing the CRUTEM3v data set, we apply a Bayesian ANOVA technique (ref. 28; code available at http://www.ncdc.noaa.gov/paleo/softlib/softlib.html) to estimate and remove means with respect to the entire 1850–2011 interval spanned by the instrumental data set.

The maximum latewood-density data set[32,33] is on the same spatial grid as the instrumental data set, and we use only the 96 grid boxes with centres poleward of 45° N. Data files and descriptions are available at http://www.cru.uea.ac.uk/~timo/datapages/mxdtrw.htm. As with the instrumental data, we apply the Bayesian ANOVA technique[28] to estimate and remove means with respect to the entire interval spanned by the data set, in this case 1400–1994.

All varve thickness records publicly available from the NOAA Paleolimnology Data Archive (http://www.ncdc.noaa.gov/paleo/paleolim/paleolim_data.html) as of January 2012 are incorporated, provided they meet the following criteria: extend back at least 200 years, are at annual resolution, are reported in length units, and the original publication or other references indicate or argue for a positive association with summer temperature. As is common[34], varve thicknesses are logarithmically transformed before analysis, giving distributions that are more nearly normally distributed and in agreement with the assumptions characterizing our analysis (see subsequent section). Records that are complete between 1400 and 1969 are standardized to unit variance and zero mean using the sample statistics computed over that interval. To obtain a more homogeneous normalization for records incomplete between 1400 and 1969, we scale the mean and standard deviation of each incomplete record to equal those statistics of the corresponding data points in the scaled, complete records. Finally, the Bayesian ANOVA technique[28] is used to remove means with respect to the entire 1400–2005 time span of the varve data set.

The ice-core data set consists of 14 of the 15 annually resolved $\delta^{18}O$ series used in a recent sea ice reconstruction[35]. We exclude the Mount Logan series, as the original reference[36] indicates it is a proxy for precipitation source region, not temperature. To increase the spatial coverage, we additionally use two $\delta^{18}O$ series from Svalbard (refs 37, 38; data available at ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/polar/svalbard/svalbard2005d18o.txt) and one each from Baffin Island (refs 39, 40; data file fisher_1998_baffin.ppd available at http://www.ncdc.noaa.gov/paleo/pubs/pcn/pcn-proxy.html) and Devon Island (ref. 41; data available at ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/polar/devon/d7273del_1yr.txt), all with annual observations. The data set spans the 1400–1998 interval; the standardization procedure is the same as for the log-transformed varve series except that it uses the 1400–1974 interval for initial standardization.

Section 1 of Supplementary Information and Supplementary Tables 1 and 2 provide additional details regarding the data sources.

Inference. Bayesian hierarchical models provide a flexible framework for combining models and observations[20]. They are generally characterized by a process level that represents the structure of the system and a data level that represents the relationships between each data type and the process. Both the process and data levels contain parameters that are uncertain and whose distributions are inferred as part of the analysis. Similarly, each observation is considered uncertain; in the current analysis this includes both the proxy and the instrumental observations. The process targeted for inference is, therefore, never directly observed but must be inferred using uncertain observations and a model with parameters whose distributions must likewise be estimated. Although Bayesian hierarchical models have been proposed and used to infer past climate[12–14,42,43], the analysis presented here is, to our knowledge, the first application of such a model to infer surface temperatures as a function of space and time from a multiproxy data set.

In this analysis, we use the Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time (BARCAST)[13,14]. In our application, the process level describes the evolution of the average April–September temperature anomaly field, $\boldsymbol{T}$, as,

$$\boldsymbol{T}_t - \mu\boldsymbol{1} = \alpha(\boldsymbol{T}_{t-1} - \mu\boldsymbol{1}) + \boldsymbol{\epsilon}_t \quad (1)$$

where μ is the mean of the temperature anomaly field, α is the coefficient of a first-order autoregressive process, the subscript t indexes the year, and $\boldsymbol{1}$ is a column vector of ones. The temperature field is represented on a $5^\circ \times 5^\circ$ grid, and we consider only continental locations that are north of 45° N. The innovation vector for each year, $\boldsymbol{\epsilon}_t$, is assumed to be an independent draw from a mean-zero multivariate normal distribution with an exponentially decaying spatial correlation: $\boldsymbol{\epsilon}_t \sim N(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}_{ij} = \sigma^2 \exp(-\phi\ |\boldsymbol{s}_i - \boldsymbol{s}_j|)$, and $|\boldsymbol{s}_i - \boldsymbol{s}_j|$ is the distance between the ith and jth elements of the field vector, $\boldsymbol{T}_t$.

The data level describes the relationships between the true temperature anomalies and the instrumental and proxy observations of these anomalies. Instrumental observations, $\boldsymbol{W}_{0,t}$, are modelled as noisy versions of the true anomalies at the corresponding locations,

$$\boldsymbol{W}_{0,t} = \boldsymbol{H}_{0,t}\boldsymbol{T}_t + \boldsymbol{e}_{0,t} \quad (2)$$

The noise terms are assumed to be independent draws from a multivariate normal distribution, $\boldsymbol{e}_{0,t} \sim N(0, {\tau_0}^2\boldsymbol{I})$, where $\boldsymbol{I}$ is the identity matrix, and $\boldsymbol{H}_{0,t}$ is a selection matrix of zeros and ones that picks out elements of $\boldsymbol{T}_t$ for which there are instrumental observations.

The types of proxy observations included are tree-ring density chronologies, ice-core $\delta^{18}O$ series, and log-transformed lake-varve thickness series. Each type, k, is assumed to have a linear relationship with the local value of the true temperature,

$$\boldsymbol{W}_{k,t} = \beta_{k,1}\boldsymbol{H}_{k,T}\boldsymbol{T}_t + \beta_{k,0}\boldsymbol{1} + \boldsymbol{e}_{k,t} \quad (3)$$

where $\beta_{k,1}$ and $\beta_{k,0}$ are respectively the slope and intercept terms and $\boldsymbol{H}_{k,t}$ is, as above, a selection matrix. The noise terms are once more assumed to be independent normal draws, $\boldsymbol{e}_{k,t} \sim N(0, {\tau_k}^2\mathbf{I})$. The regression parameters vary between proxy types, but are constant for all proxies of a given type.

Prior distributions are placed on each of the parameters included in the model: α, μ, ϕ, σ^2, ${\tau^2}_{k=0\ldots3}$, $\beta_{k=1\ldots3,1}$, $\beta_{k=1\ldots3,0}$, as well as for $\boldsymbol{T}$ in the year before observations become available. Priors are selected to be proper, weakly informative, and—where possible—conjugate[20]. Bayes' rule is used to calculate the posterior distribution of the parameters and field given the observations and priors. A Gibbs sampler with a single Metropolis step (for ϕ) is then used to draw from the posterior distribution[20]. Further details of the inference are available in an earlier publication[13], code is available at http://www.ncdc.noaa.gov/paleo/softlib/softlib.html, and convergence of the Gibbs sampler is discussed in Section 3.5 of the Supplementary Information.

The result of the analysis is an ensemble of draws of both the parameters and the temperature anomalies, each of which is equally likely given the data, priors and modelling assumptions. Furthermore, each ensemble member will have variability similar to the actual temperature anomalies[14], insomuch as the model and data are correctly represented. The median across the ensemble (Fig 1a, b; Supplementary Fig. 5), however, has attenuated variability[14], especially in data-poor parts of the reconstruction. This attenuation provides a more accurate estimate of the past temperature, though not the variability of that temperature, and is generally used as our best estimate.

For the purposes of comparing the climatic information content of the different proxies (for example, Fig. 2c), it is possible to run BARCAST in a reduced mode, using a subset of the data to update the temperature field while sampling all model parameters from the posterior distributions resulting from the analysis with the full data set. That is, at each iteration of the Gibbs sampler, a vector of parameters is drawn from the posterior distribution derived using all data, and then the draw of the temperature field is updated using these parameters and a given subset of the data. It is likewise possible to simulate the natural variability of the temperature field (for example, Fig. 3) by not applying any of the data constraints. In this case, the process level model has parameters constrained by the data but the specific evolution of the temperature field is unconstrained. Comparison of the variability between the constrained and unconstrained simulations indicates the extent to which the data controls the solution (see Supplementary Information Sections 2.5 and 6).

Assumptions and implications. The stationary, isotropic and exponentially decaying spatial covariance model specified for the temperature anomalies is a simplifying assumption that does not account for directionality and long-range covariance relationships in the climate system. Indeed, many palaeoclimate reconstruction techniques, generally based on eigendecompositions of sample covariance or correlation matrices, are explicitly designed to exploit such covariance patterns[9,44–46]. This class of methods assumes that the characteristic spatial structures identified in the calibration interval are constant in time, but have varying amplitudes; results can be strongly dependent on the particulars of how these modes of variability are determined and used[47]. BARCAST, in contrast, relies on the simpler assumption of a temporally constant decorrelation length scale and represents each observation as indicative of local temperature. This local representation (equation (3)) is similar to the proxy representation in the LOC method[18,19,48], although BARCAST additionally models errors in the instrumental observations and arrives at an estimate of the spatial field of temperature anomalies[49,50].

Temporal stationarity is assumed through the first-order autoregressive description of interannual temperature variability (equation (1)). This parameterization lacks any representation of long-term temperature trends, or associations between temperatures and climate forcings[43,51]. In some sense, however, this is advantageous as the process level is agnostic regarding any changes, and inferences concerning the unprecedented nature of recent extremes are consequently derived exclusively from the data. Furthermore, this approach permits exploration of exactly where the assumption of strict temporal stationarity fails (see Fig. 3).

The assumption that the space–time covariance function is separable—that is, that it can be factored into the product of purely spatial and purely temporal elements—is also unlikely to hold in detail[12]. Predictive performance, however, is often not affected by incorrectly assuming a separable covariance form[52]. With regard to both stationarity and separability, BARCAST has been shown to perform at a level comparable to or better than other climate field reconstruction techniques[22], even when the underlying assumptions are not strictly met by the data[14]. The validation metrics reported in Supplementary Information Section 5 also suggest that BARCAST provides an adequate statistical description of the data for the present purposes.

Analysis of results. The analysis presented here is based on 4,000 posterior draws taken from four parallel Gibbs samplers. Spatial mean values are computed by weighting the grid boxes by land area according to a 0.25° land–sea mask[31]. Taking the 50th percentile of the 4,000-member ensemble at each year of the spatial mean time series results in a best estimate of the time series, while taking the 5th and 95th percentiles produces 90% pointwise credible intervals[20], which are used to indicate the uncertainty in the reconstruction at each year.

Pathwise 90% credible intervals are calculated by inflating the pointwise intervals such that the envelope contains 90% of the posterior time series in their entirety[15,17,29]. The two uncertainty intervals have different interpretations, with the pointwise intervals covering the true temperature anomaly for 90% of the years, whereas the time series of true temperature anomalies lies entirely within the pathwise envelope with 90% probability. The statement that certain recent years are warmer than all previous years with $P > 0.95$ follows from noting that the 90% pathwise uncertainties for these years lie entirely above those for all years in the 1400–2004 interval. Note that the statement holds at $P > 0.95$ because the test is one sided, as we have prior reason to believe that recent years are warmer than usual. Uncertainties for temporally smoothed time series are estimated in the same manner, after first smoothing each ensemble member using a nine-point Hanning window. Point estimates and uncertainties in the centennial-scale trends are derived by calculating the trend for each ensemble member at each position of a sliding 100-year window and then calculating both pointwise and pathwise uncertainties.

Probability estimates corresponding to specific statements, such as the probability that western Russian temperatures achieved a maximum in 2010, are obtained by calculating the proportion of ensemble members for which the statement is true. Performing such an analysis at each location leads to the maps shown in Fig. 2a. Calculating the fraction of locations for which each year is warmest or coldest, averaging results across the ensemble members and binning the years by decades results in the histograms shown Fig. 2b. Figure 2c is derived in the same manner, but from a 4,000 member ensemble obtained by running BARCAST in reduced mode using only the ice-core and log-transformed lake-varve series.

To assess recent extremes, we simulate temperature anomalies and instrument-like observations over the past 20 years using median parameter values, and in the latter case additionally record the maxima at each location as well as the five largest values in both space and time. Distributions of the simulated instrumental quantities are built up by repeating the procedure 10,000 times. When shifting the mean of these simulations to match that inferred over the past 20 years (Fig. 3), we use the simple average across the locations, ensemble members, and years within the 1992–2011 interval. Note that this mean differs slightly from the mean of the spatial average time series because of the spatial weighting inherent in the latter. Figure 3b shows the single instrumental simulation that is closest to the actual instrumental observations according to the variance of the site-wise maxima. Additional results and model diagnostics are available in Supplementary Information Sections 2 and 3 and a more detailed discussion on inferring extremes in the presence of uncertainty in Supplementary Information Section 6.

Robustness. To examine the robustness of our results to specific data types, we run BARCAST in five different reduced mode formulations using the following subsets of data types: tree-ring densities alone, ice-core series alone, lake-varve series alone, ice-core and lake-varve series together, and instrumental series alone. Results of the proxy-only analyses are then compared with each other, to the main analysis, to the instrumental-only analysis, and to the withheld instrumental observations. The time series of spatial-average temperature and centennial-scale slopes, as well as the distribution of extremes for the spatial mean, for centennial trends in the spatial mean, and as a function of space, are all compared with each other. For the spatial average, we also consider the correlations and root-mean-square error between each proxy-only analysis and the instrumental-only analysis over three different time intervals: 1850–1959, 1850–1994 and 1960–94. In general, we find that the proxy-only analyses provide consistent inferences with one another and with the instrumental-only predictions, when accounting for the uncertainties in each analysis (see Supplementary Information Sections 5.1–5.4).

An important exception is that the predictions from the tree-ring densities alone do not track the warming seen in all other data sets in the latter half of the twentieth century, a finding consistent with the so-called divergence problem[33,53]. To assess the robustness of our results to this divergence, we re-ran the full analysis excluding the post-1960 tree-ring-density observations and found no qualitative change in results (see Supplementary Information Section 4). The primary reason for consistency between analyses which include and exclude the post-1960 tree-ring-density observations is that instrumental data are of sufficient quality and number post-1960 so as to dominate the solution irrespective of the tree-ring data. It is also the case that the parameterization of the tree-ring-density relationship with temperature is primarily constrained by its relationship with data before 1960.

To assess the variability in the ensemble of posterior draws, we simulate the withheld instrumental observations in each proxy-only analysis and examine rank verification histograms[54] for the withheld instrumental observations. To assess the coverage rates of the credible intervals, we calculate the percentage of the withheld instrumental observations that fall within the nominal 90% intervals. Results show that the ensembles of predictions of the withheld instrumental values generally have about the correct variability, and that the actual coverage rates are generally within 10% of the nominal rate. An exception is, again, for the tree-ring-density analysis during the post-1960 interval, where the coverage rate is about 15% too low and the shape of the rank verification histogram is indicative of a low bias. See Supplementary Information Section 5.5 for further details.

31. Rodell, M. *et al.* The global land data assimilation system. *Bull. Am. Meteorol. Soc.* **85,** 381–394 (2004).
32. Briffa, K. *et al.* Tree-ring width and density data around the northern hemisphere: part 1, local and regional climate signals. *Holocene* **12,** 737–757 (2002).
33. Briffa, K. *et al.* Tree-ring width and density data around the northern hemisphere: part 2, spatio-temporal variability and associated climate patterns. *Holocene* **12,** 759–789 (2002).
34. Loso, M. Summer temperatures during the medieval warm period and little ice age inferred from varved proglacial lake sediments in southern Alaska. *J. Paleolimnol.* **41,** 117–128 (2009).
35. Kinnard, C. *et al.* Reconstructed changes in Arctic sea ice over the past 1,450 years. *Nature* **479,** 509–512 (2011).
36. Fisher, D. *et al.* The Mt Logan Holocene–late Wisconsinan isotope record: tropical Pacific–Yukon connections. *Holocene* **18,** 667–677 (2008).
37. Isaksson, E. *et al.* Climate oscillations as recorded in Svalbard ice core δ^{18}O records between 1200–1997 AD. *Geogr. Ann. A* **87,** 203–214 (2005).
38. Isaksson, E. *et al. Svalbard Ice Cores 600 Year Annual δ^{18}O Data* (IGBP PAGES/World Data Center for Paleoclimatology, Data Contribution Series no. 2011-068. NOAA/NCDC Paleoclimatology Program, Boulder 2011); available online at ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/polar/svalbard/svalbard2005d18o.txt.
39. Fisher, D. *et al.* Penny ice cap cores, Baffin Island, Canada, and the Wisconsinan Foxe Dome connection: two states of Hudson Bay ice cover. *Science* **279,** 692–695 (1998).
40. Mann, M. E. *et al.* Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc. Natl Acad. Sci. USA* **105,** 13252–13257 (2008).
41. Fisher, D. Comparison of 105 years of oxygen isotope and insoluble impurity profiles from the Devon Island and Camp Century ice cores. *Quat. Res.* **11,** 299–305 (1979).
42. Haslett, J. *et al.* Bayesian palaeoclimate reconstruction. *J. R. Stat. Soc. A* **169,** 395–438 (2006).
43. Li, B., Nychka, D. & Ammann, C. The value of multi-proxy reconstruction of past climate. *J. Am. Stat. Assoc.* **105,** 883–895 (2010).
44. Mann, M., Bradley, R. & Hughes, M. Global-scale temperature patterns and climate forcing over the past six centuries. *Nature* **392,** 779–787 (1998).
45. Schneider, T. Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. *J. Clim.* **14,** 853–871 (2001).
46. Smerdon, J., Kaplan, A., Chang, D. & Evans, M. A pseudoproxy evaluation of the CCA and RegEM methods for reconstructing climate fields of the last millennium. *J. Clim.* **23,** 4856–4880 (2010).
47. Christiansen, B., Schmith, T. & Thejll, P. A surrogate ensemble study of sea level reconstructions. *J. Clim.* **23,** 4306–4326 (2010).
48. Christiansen, B. Reconstructing the NH mean temperature: can underestimation of trends and variability be avoided? *J. Clim.* **24,** 674–692 (2011).

49. Tingley, M. & Li, B. Comments on 'Reconstructing the NH mean temperature: can underestimation of trends and variability be avoided?'. *J. Clim.* **25,** 3441–3446 (2012).
50. Christiansen, B. Reply to ''Comments on 'Reconstructing the NH mean temperature: can underestimation of trends and variability be avoided?'''. *J. Clim.* **25,** 3447–3452 (2012).
51. Lee, T., Zwiers, F. & Tsao, M. Evaluation of proxy-based millennial reconstruction methods. *Clim. Dyn.* **31,** 263–281 (2008).
52. Genton, M. Separable approximations of space–time covariance matrices. *Environmetrics* **18,** 681–695 (2007).
53. D'Arrigo, R., Wilson, R., Liepert, B. & Cherubini, P. On the 'Divergence problem' in northern forests: a review of the tree-ring evidence and possible causes. *Glob. Planet. Change* **60,** 289–305 (2008).
54. Hamill, T. Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Weath. Rev.* **129,** 550–560 (2001).