Reconstructing Past Climate from Noisy Data
Hans von Storch et al.
Science 306, 679 (2004);
DOI: 10.1126/science.1096109

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of September 4, 2014):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/content/306/5696/679.full.html

Supporting Online Material can be found at:
http://www.sciencemag.org/content/suppl/2004/11/12/1096109.DC1.html

A list of selected additional articles on the Science Web sites related to this article can be found at:
http://www.sciencemag.org/content/306/5696/679.full.html#related

This article cites 22 articles, 5 of which can be accessed free:
http://www.sciencemag.org/content/306/5696/679.full.html#ref-list-1

This article has been cited by 133 article(s) on the ISI Web of Science

This article has been cited by 16 articles hosted by HighWire Press; see:
http://www.sciencemag.org/content/306/5696/679.full.html#related-urls

This article appears in the following subject collections:
Atmospheric Science
http://www.sciencemag.org/cgi/collection/atmos
Reconstructing Past Climate from Noisy Data

Hans von Storch,¹,6 Eduardo Zorita,¹ Julie M. Jones,¹ Yegor Dmitriev,¹ Fidel González-Rouco,² Simon F. B. Tett³

Empirical reconstructions of the Northern Hemisphere (NH) temperature in the past millennium based on multiproxy records depict small-amplitude variations followed by a clear warming trend in the past two centuries. We use a coupled atmosphere-ocean model simulation of the past 1000 years as a surrogate climate to test the skill of these methods, particularly at multi-decadal and centennial time scales. Idealized proxy records are represented by simulated grid-point temperature, degraded with statistical noise. The centennial variability of the NH temperature is underestimated by the regression-based methods applied here, suggesting that past variations may have been at least a factor of 2 larger than indicated by empirical reconstructions.

Reconstruction of past climate from palaeoclimate proxy data is important for the detection of anthropogenic climate change. A number of studies have attempted to reconstruct variations in global or NH temperature within the past millennium by regressing proxy indicators and early instrumental time series on recent instrumental climate variables with high spatial resolution (1–3). Regression models are developed during the period of common instrumental and proxy data and are then applied to longer proxy records to reconstruct past climates. Similar methods have been applied for the reconstruction of atmospheric circulation indices, such as the North Atlantic Oscillation (4) or the Antarctic Oscillation (5). A number of reconstructions show that the temperatures in the last millennium were characterized by geographically varying warm values in the 11th and 12th centuries, followed by a secular cooling trend punctuated by decadal-scale colder periods in the mid-16th, early 17th, and early 19th centuries (6). These cooler intervals were followed by the marked warming experienced until today. Although the amplitude of these preindustrial variations is still debated, according to the most quoted NH temperature reconstruction [Mann, Bradley, Hughes, 1998 (MBH98)] and Mann, Bradley, Hughes, 1998 (MBH98) (1) and Mann, Bradley, Hughes, 1999 (MBH99) (2) and the most recent Intergovernmental Panel on Climate Change (IPCC) report (7), these variations were of small amplitude. However, recent studies with general circulation models suggest that these centennial variations may have been larger (8–10). We used a coupled atmosphere-ocean model simulation of the past 1000 years as surrogate climate to test whether

References and Notes

33. Accessing the lower wavelength region of the bending modes (−1500 cm−1) required modification of our pulsed infrared laser (LaserVision) by addition of a third nonlinear mixing stage based on an AgGaSe2 crystal. This arrangement provides about 100 μJ of energy in the bending region.
34. Because of the near mass degeneracy between Ar and (D2O)4−, we verified that the dimer and trimer clusters exhibit no photofragmentation or sharp bands, consistent with their low VDEs. Thus, the (D2O)4−−Ar+ spectrum could be selectively monitored by detecting the photofragment anions that arise only from this species.
43. We thank the U.S. Department of Energy (grant DE-FG02-00ER15066) for support of this research direction, as well as NSF, which provided the equipment used in this study. We also thank K. D. Jordan for valuable discussions regarding the formation mechanisms at work in these clusters and T. Zwier and J. Stearns for help with the wavelength extension of the laser.

15 July 2004; accepted 20 August 2004
the reconstruction method of MBH98 and MBH99 and a much simpler regression method can yield realistic estimates of the multi-decadal and secular temperature variations. A number of modeling studies of the evolution of the climate in the past centuries (11–14) pose some questions about the reliability of empirical reconstructions based on regression methods. For instance, concerning the cooling around 1700, the reconstructions by MBH98 have agreed with the results obtained with the the Goddard Institute for Space Studies model only in the model version with a low climate sensitivity (0.4K/W/m$^2$) (10). Similarly, the agreement between an energy-balance model (11) and reconstructions has been achieved by prescribing a model sensitivity to changes in radiative forcing of 0.5K/W/m$^2$. These values of climate sensitivity are at the low end of the range of the models included in the IPCC analysis (7). Other reconstructions that indicate markedly stronger cooling in the 16th to 18th centuries are, for instance, the result of empirical methods that explicitly aim to preserve low-frequency variability (15) and the borehole-based reconstructions (16). Interestingly, the latter is not based on empirical regression methods. This apparent discrepancy poses a question as to whether model simulations overestimate secular climate variability or regression-based reconstructions underestimate it.

The reliability of these empirical methods at centennial time scales can be tested in the surrogate climate simulated by three-dimensional climate models driven by plausibly estimated historical external forcing. The results of the reconstruction process can then be validated against the climate fields simulated by the climate model. The data representing the proxy records are climate variables simulated at grid-box resolution, which can be degraded with statistical noise to mimic more realistic data [so-called pseudoproxies (17)]. Here, we follow this strategy with the output of a climate simulation of the past millennium with the coupled atmosphere-ocean general circulation model European Centre Hamburg 4–Hamburg Ocean Primitive Equation–G (ECHO-G) (18), driven by estimations of historical climate forcing. This simulation provides a data set in which the potential nonstationarity of the covariances and the length of the time series are similar to those found in applications of the empirical reconstruction methods.

This simulation reproduces warming around 1100 and extended coolings over the Spörer, Maunder, and Dalton Minima as near-global events, as well as the recent anthropogenic warming. Compared with the reconstructions of MBH99, however, the variations are stronger. For the purpose of this paper, it is not critical if the simulation is not absolutely realistic because of model limitations (e.g., coarse resolution or deficient representation of processes) or uncertainties in external forcings. The crucial point is that the model simulates a reasonable, internally consistent climate, and the external forcing lies within the envelope of possible values. In this case, it will be used as a virtual world to determine the skill of regression-based reconstruction methods like MBH99 to estimate its temperature variations.

Here, we focus on the reconstruction of NH temperature. For this analysis, we applied as realistically as possible the statistical method of MBH98. However, when arbitrary choices were required, the a priori most favorable for the statistical method were implemented, thus probably minimizing the loss of variance in the statistical reconstructions. For instance, the proxy network in MBH98 gets coarser backward in time, but the pseudoproxy network in this study was not decimated to avoid loss of skill. Also, only temperature pseudoproxies were used to reconstruct the temperature evolution. The pseudoproxies were generated by adding a statistical white noise to the simulated temperatures in grid points colocated to the MBH98 proxy network (1). Several tests with varying amounts of noise were carried out.

The loss of variance resulting from a reconstruction-type method may be simply conceptualized (18). The proxy data $P$ are thought to blend local temperature $T_l$ and unrelated variability $\varepsilon$: $P = aT_l + \varepsilon$. The temperature variations are essentially estimated as $T^* = \beta P$, where $\beta = \rho \sigma_p/\sigma_p < 1$ and $\rho$ is the correlation between $T_l$ and $P$. Therefore, $\text{Var}(T^*) = \rho^2 \text{Var}(T) < \text{Var}(T_l)$. In case of proxy data, the correlation $\rho$ is mostly on the order of 0.4 to 0.7 (8), resulting in a leakage or variance on the order of 50 to 80%. In particular, if $T_l$ has a red spectrum and $\varepsilon$ has a white spectrum, $T^*$ will underestimate the low-frequency variance of $T_l$. In the case of MBH98 and other reconstructions, the methodological process is more sophisticated, but the fundamental problem of the loss of variance resulting from noisy proxy data may also exist in these studies (19). This loss of variance, also known in areas such as regionalization and long-term forecasting (20), is sometimes accounted for by artificially inflating the parameter $\rho$. For paleoclimate reconstructions, $\rho$ should be made time-scale dependent, and this dependency is unknown.

To implement the method of MBH98, we selected model grid boxes collocated with their proxy data network (Fig. 1, red pixels) and added white noise (21) to the grid-point temperatures $T_g$, so that the pseudoproxy data are $P = T_g + \varepsilon$. The variance of $\varepsilon$ varies between $m = 0$ and $m = 4 \times \text{Var}(T_g)$. The correlation between the $T_g$ and $P$ is then $(1 + m)^{-1/2}$. Thus, with $m = 0$, the local $P$ variance described by $T_g$ is 100%; for $m = 1$, when noise with the same variance as that of the local temperature is added, the percentage of described variance is 50%. For $m = 4$, the described variance is only 20%. Ideally, the reconstructions would coincide with the simulated NH temperature, but actually they do not, even for $m = 0$ (Fig. 2A, illustrating the loss of variance induced by the method alone). The short-term variations are reasonably reproduced, at least for $m < 4$. For instance, on an interannual time scale, the fit between simulated and reconstructed NH temperature is good, with a calibration reduction-of-error statistics of 0.7 for perfect pseudoproxies and 0.30 for pseudoproxies with $\varepsilon = 0.5$. The substantial underestimation of low-frequency temperature variations is evident from Fig. 2B. For example, only 20% of the 100-year variability is recovered when the noise level is 50%. For time scales of 20 years, about three-quarters of the variability is lost. Similar results are obtained with a simulation with the third Hadley Centre coupled model (HadCM3), demonstrating that the results obtained here are not dependent on the particular climate characteristics of the ECHO-G simulation (18). Also a spatially varying level of noise does not essentially modify these conclusions (18).

Our setup allowed the test of a number of hypotheses. The first hypothesis is that the inclusion of more instrumental data would improve the estimate, as the multiproxy data used by MBH99 contained a number of long instrumental temperature data, which start...
typically at the end of the 18th century. To test the influence of such instrumental data, we included grid-box temperatures without adding noise (18). The effect of this modification on the hemispheric temperature was small; the differences in the reconstructed temperature anomalies were within a range of 2%. This can be explained by the relatively low number of perfect pseudoproxies included in comparison with the total number of pseudoproxies and by the built-in robustness against local influences of the inverse regression method used by MBH98, because the signal is extracted nonlocally from the whole proxy data set simultaneously (18). Other direct regression methods, aimed at more regionally limited temperature reconstructions, do show an improvement when instrumental records are included in the proxy network (22).

The second hypothesis is related to the sparseness of the proxy locations (Fig. 1, red pixels). The proxy data set was enlarged by adding 18 locations in Asia and Africa (Fig. 1, blue pixels) to increase the spatial coverage. This leads to a minor improvement in the NH temperature reconstruction (Fig. 3), which was largely independent of the inclusion or exclusion of Southern Hemisphere pseudoproxies.

Finally, we tested whether the range of variability present in the instrumental period is sufficient to reconstruct the climate of past centuries. To test this hypothesis, 40 years were taken from the Late Maunder Minimum (1680 to 1720) and 40 years from the early part of the 20th century (1900 to 1940) to calibrate the statistical model, thereby expanding the range of temperature variability present in the pseudoproxies. When the proxies are free of noise, the reconstruction of the simulated NH temperature is greatly improved (Fig. 3). With 50% local noise included, the reconstruction is also improved, although the loss of low-frequency variance is still large. Therefore, augmenting the variability in the calibration period improves the skill, but obviously this is limited by the available observational record.

A further question is whether the limitations we have found are common to regression methods in general. Thus, two further approaches were tested. In the first, local temperatures were estimated by a linear regression from pseudoproxies, and the local temperature estimations were spatially averaged to derive the NH temperature. This method mimics the situation in which, for example, local dendrochronologies calibrated in terms of local temperature are just arithmetically averaged. In the second approach, the pseudoproxies at the various locations were directly simply averaged. This is more similar to the borehole methodology (16). For the first method, we found qualitatively similar but quantitatively even worse problems than with the MBH98 method; that is, the underestimation of low-frequency variability for a given amount of noise is greater than for MBH98, whereas the second method returns good estimates of NH temperature, with very little loss of variance. With 75% variance noise (Fig. 4), this result is not surprising given that the first method suffers from the variance loss related to regression, whereas in the second, the noise contributions are simply averaged out.

Hints of the underestimation of low-frequency variability by empirical reconstruction methods have been found in previous studies, based either on short data sets (17) or climate simulations with fixed external forcing (23). In a study based entirely on an instrumental data set (17), the spectrum of the difference between the reconstructed and observed global mean annual temperature is, albeit consistent with a white noise assumption, slightly red. In a further analysis of an instrumental data set and data from a long control simulation with the Geophysical Fluid Dynamics Laboratory climate model (with constant external forcing) and a relatively short simulation of 143 years driven by varying external forcing (23), the spectra of the temperature differences from the analysis of control simulation are red (although again statistically compatible with white noise assumption). In this externally forced simulation, it was found that the temperature reconstructions are biased if the external forcing leads to nonstationary behavior in the verification period. In a long control simulation (1000 years) with the model ECHO-G (24), the spectrum of the reconstructed annual global temperature underestimates the spectrum of the simulated global temperature at very low frequencies.

Fig. 2. (A) The NH annual temperature evolution over the past 1000 years. The NH annual temperature simulated by the model ECHO-G and MBH98 reconstructions of this temperature from 105 model grid points mimicking the multiproxy network of MBH98. Increasing amounts of noise have been added to the grid-point temperatures to mimic the presence of other than temperature signals in the proxies. The corresponding local correlation is also indicated. The 2σ uncertainty range (derived as in MBH98 from the variance of the interannual residuals) for the different noise levels is indicated. The reconstruction with $p = 0.5$ is shown with its 2σ uncertainty range. (B) The spectra of the NH annual temperatures shown in (A).
Climate simulations of the past millennium are burdened by model limitations and uncertainties in the external forcing, and therefore their output must be considered with care. However, they provide a surrogate climate realistic enough to conclude that the use of the regression methods considered here, which exploit short-term cross-correlations to reconstruct past climates, suffer from marked losses of centennial and multidecadal variations. This conclusion probably applies to most other regression-based methods. Other methods that estimate past temperatures with physical, as opposed to statistical, methods (e.g., borehole temperature profiles (16)), or regression methods that address retention of the information of the low-frequency variability contained in the proxy indicators (25), may be in theory free from this specific caveat. Our results indicate that a detailed testing of these reconstruction methods in simulated climates should be an essential part in the reconstruction process and may help in the design of better reconstruction methods.

**References and Notes**

18. Materials and Methods are available as supporting material on Science Online.
19. For instance, the standard deviation of the MBH99 reconstructed NH temperature in the period of 1948 to 1980 is 0.9K, compared with 1.4K derived from the National Centers for Environmental Protection reanalysis or from the Jones et al. (26) NH gridded instrumental data.
21. Local noise could in many cases be expected to be either white or red. In any case, red noise degrades the reconstructions, as in (17).
27. This work was carried out within the projects German Climate Research Program [DEXUM], [German Federal Ministry for Education and Research (BMBF)]; Simulations, Observations, and Palaeoclimate data (SOAP) [European Union, EVK2-CT-2002-00160]; and REN-2000-0786CLI [Comisión Interministerial de Ciencia y Tecnología]. S.F.B.T. was supported by a UK Government Meteorological Research contract and SOAP. Computer time for HadCM3 simulations was funded by Defra under contract PEC/7/112. Data distribution was through SOAP. Three anonymous reviewers greatly contributed to the improvement of the original manuscript. We acknowledge fruitful suggestions by T. Stocker.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1096109/DC1
SOM Text
Figs. S1 and S2
References
27 January 2004; accepted 23 August 2004
Published online 30 September 2004; 10.1126/science.1096109
Include this information when citing this paper.