Sciencexpress Report

Reconstructing Past Climate from Noisy Data

Hans von Storch,[1*] Eduardo Zorita,[1] Julie Jones,[1] Yegor Dimitriev,[1] Fidel González-Rouco,[2] Simon Tett[3]

[1]Institute for Coastal Research, GKSS Research Centre, Geesthacht Germany. [2]Department of Astrophysics and Atmospheric Physics, Universidad Complutense, Madrid, Spain. [3]UK Meteorological Office, Hadley Centre (Reading Unit), Meteorology Building, University of Reading, Reading, RG6 6BB.

*To whom correspondence should be addressed. E-mail: storch@gkss.de

Empirical reconstructions of the Northern Hemisphere (NH) temperature in the last millennium based on multy proxy records depict small-amplitude variations followed by a clear warming trend in the last two centuries. We use a coupled atmosphere-ocean model simulation of the last 1000 years as a surrogate climate to test the skill of these methods, particularly at multidecadal and centennial timescales. 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 two larger than indicated by empirical reconstructions.

Reconstruction of past climate from palaeoclimate proxy data is important for detection of anthropogenic climate change. A number of studies have attempted to reconstruct variations in global or Northern Hemisphere (NH) temperature within the last 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 are applied for the reconstruction of atmospheric circulation indices, e.g. 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, respectively (*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 (1, 2, hereafter MBH98 and MBH99) and the last 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*). In the following we will test, using a coupled atmosphere-ocean model simulation of the past 1000 years as surrogate climate, whether the reconstruction method of MBH98 and MBH99 and a much simpler regression method can yield realistic estimates of the multidecadal and secular temperature variations.

A number of modeling studies of the evolution of the climate in the last 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 agree with the results obtained with the GISS model only in the model version with a low climate sensitivity (0.4 $K/W/m^2$) (*10*). Similarly, the agreement between an energy balance model (*11*) and reconstructions is achieved by prescribing a model sensitivity to changes in radiative forcing of 0.5$K/W/m^2$. 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-18th are, for instance, the result of empirical methods that explicitly aim to preserve low-frequency variability (*15*) and the borehole-based reconstructions (*16*), which, interestingly, is not based on empirical regression methods. This apparent discrepancy poses the question as to whether model simulations overestimate secular climate variability or regression-based reconstructions underestimate it.

The reliability of these empirical methods at centennial timescales 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, that can be degraded with statistical noise to mimic more realistic data [so-called pseudo-proxies (*17*)]. Here, we follow this strategy using the output of a climate simulation of the last millennium with the coupled AO-GCM ECHO-G (*18*), driven by estimations of historical climate forcing. This simulation provides a data set, where 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 to the reconstructions of MBH99, the variations are, however, stronger. For the purpose of this paper it is not critical if the simulation is not absolutely realistic due to model limitations (e.g. coarse resolution or deficient representation of processes) or uncertainties in external. 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 MBH98 to estimate its temperature variations.

Here, we focus on the reconstruction of Northern Hemisphere (NH) temperature. For this analysis we apply as realistically as possible the statistical method of MBH98. However, when arbitrary choices are required, the a priori most favorable for the statistical method are implemented, thus probably minimizing the loss of variance in the statistical reconstructions. For instance, the proxy network in MBH98 gets coarser backwards in time, but the pseudo-proxy network in this study is not decimated to avoid loss of skill. Also, only temperature pseudo-proxies are used to reconstruct the temperature evolution. The pseudo-proxies are generated by adding a statistical white noise to the simulated temperatures in grid points co-located to the MBH98 proxy network (*1*). Several tests with varying amounts of noise have been carried out.

The loss of variance due to a regression-type method may be simply conceptualized (*18*). The proxy data P are thought to blend local temperature T_l and unrelated variability ε: $P = \alpha T_l + \varepsilon$. The temperature variations are essentially estimated as $T_l^* = \beta P$ where $\beta = \rho\sigma_T/\sigma_P < 1$ and ρ is the correlation between T_l and P. Therefore Variance(T^*) = ρ^2Variance(T_l) < Variance(T_l). In case of proxy data, the correlation ρ is mostly of the order of 0.4-0.7 (*8*), resulting in a leakage or variance of the order of 50-80%. In particular, if T_l has a red spectrum and ε 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 due to noisy proxy data may exist also in these studies (*19*). This loss of variance, also known in areas such as regionalization and long-term forecasting (*20*), is sometimes parched by artificially inflating the parameter ρ. For paleoclimate reconstructions, ρ should be made timescale dependent, and this dependency is unknown.

To implement the method of MBH98, we select model grid boxes co-located with their proxy data network (Fig. 1; red pixels), and add white noise (*21*) to the grid-point temperatures T_g, so that the pseudo-proxy data are $P = T_g + \varepsilon$. The variance of ε varies between m = 0 and m = 4 ×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 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 interannual timescale the fit between simulated and reconstructed NH temperature is good, with a calibration Reduction of Error statistics of 0.7 for perfect pseudo-proxies and 0.30 for pseudo-proxies with ε=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 HadCM3 climate model, 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 set-up allows the test of a number of hypotheses. The first hypothesis is that the inclusion of more instrumental data would improve the estimate, as the multi-proxy data used by MBH99 contained a number of long instrumental temperature data, which start typically at the end of 18th century. To test the influence of such instrumental data, we have 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 pseudo-proxies included in comparison with the total number of pseudo-proxies, and by the built-in robustness against local influences of the inverse regression method used by MBH98, since the signal is extracted non-locally 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 net-work (*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 15 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 whether Southern Hemisphere pseudo-proxies were included.

Lastly, we test 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 (1680-1720) and 40 years from the early part of the 20th century (1900-1940) to calibrate the statistical model, thereby expanding the range of temperature variability present in the pseudo-proxies. 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 pseudo-proxies, and the local temperature estimations were spatially averaged to derive the NH temperature. This method mimics the situation in which e.g. 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 find qualitatively the same, but quantitatively even worse problems than with the MBH98 method, i.e. 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 as the first method suffers from the variance loss related to regression, while 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 GFDL 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.

Climate simulations of the last millennium are burdened by model limitations and uncertainties in the external forcing, and therefore their output has to 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 multi-decadal variations. This conclusion probably applies to most other regression-based methods. Other methods that estimate past temperatures using 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

1. M.E. Mann, R. S. Bradley, M.K. Hughes, *Nature* **392**, 779 (1998).
2. M.E. Mann, R. S. Bradley, M.K. Hughes *Geophys. Res. Lett.*, **26**, 759 (1999).
3. P.D.Jones, K.R.Briffa, T.P. Barnett, S.F.B.Tett, *The Holocene* **8,** 455 (1998).
4. J. Luterbacher, C. Schmutz, D. Gyalistras, E. Xoplaki, H. Wanner, *Geophys Res. Lett.,* **26**, 2745 (1999).
5. J.M. Jones, M. Widmann, *J. Clim.* **16**, 3511 (2003).
6. P. D. Jones, T. J. Osborn, K. R. Briffa, *Science* **292**, 662 (2001).
7. Houghton *et al.* , *Climate Change 2001: The scientific basis*. (Cambridge University Press, 2001).
8. P.D. Jones, M.E. Mann, *Rev. Geophys*. **42**, RG2002 (2004).
9. E. Zorita *et al.* , *Meteorol. Z.*, in press. (2004).
10. D. Rind *et al.* , *J. Climate* **17**, 906 (2004).
11. T.J Crowley, *Science* **289**, 270 (2000).
12. Gerber S., F. Joos, P. Breuger, T.F. Stocker, M.E. Mann, S. Stich, M. Scholze*, Clim. Dyn.* **20**, 281-299 (2003).
13. H. Goosse *et al.* , *Geophys. Res. Lett.* **31** (6): Art. No. L06203 (2004).
14. E. Bauer, M. Claussen, V. Brovkin, A. Huenerbein. *Geophys. Res. Lett.* **30,** art no**.** 1276 (2003).
15. J. Esper, E.R.Cook, F.H. Schweingruber, *Science* **295**, 2250 (2002).

16. S. H. Huang, H.N. Pollack, P.Y. Shen, *Nature* **403**, 756 (2000).
17. M.E. Mann, S. Rutherford,: *Geophys. Res. Lett*. **29**, 139-1 (2002).
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 1948-1980 is 0.9 K, compared to 1.4K derived from the NCEP reanalysis or from the Jones et al (1998) NH gridded instrumental data.
20. H. von Storch, *J. Clim*., **12**, 3505 (1999).
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*).
22. J. Luterbacher, D. Dietrich, E. Xoplaki, M. Grosjean, H Wanner, *Science* **203**, 1499 (2004).
23. S. Rutherford, M. E. Mann, T.L. Delworth, R.J. Stouffer,: *J. Clim*. **16**, 462. (2003).
24. E. Zorita, F. González-Rouco, S. Legutke,: *J. Clim.,* **16**, 1378 (2003).
25. T. J. Osborn, K. R. Briffa, *Dendrochronologia* **18**, 9 (2000).
26. This work was carried out within the projects DEKLIM (German BMBF), SO&P (EU, EVK2-CT-2002-00160) and REN-2000-0786cli (Spanish CICYT). SFBT was supported by UK GMR contract and SOAP. Computer time for HadCM3 simulations was funded by Defra under contract PECD/7/12. Data distribution through SO&P. 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

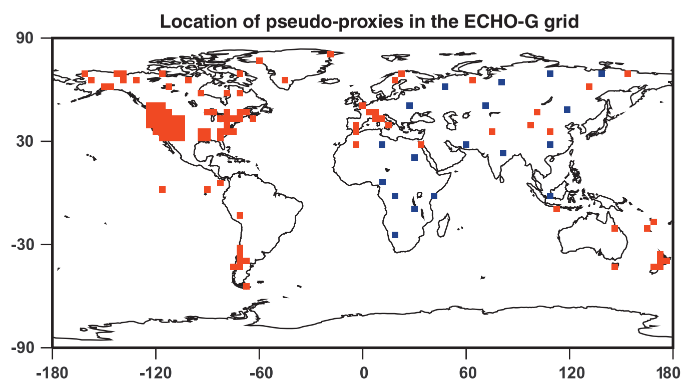
27 January 2004; accepted 23 August 2004
Published online 30 September 2004;
10.1126/science.1096109
Include this information when citing this paper.

Fig. 1: Grid-boxes in the ECHO-G model, from which simulated temperatures are used to estimate Northern Hemisphere temperatures. The red pixels are used for the basic reconstruction; the blue pixels are added in a test of whether a better spatial coverage would improve the reconstruction.

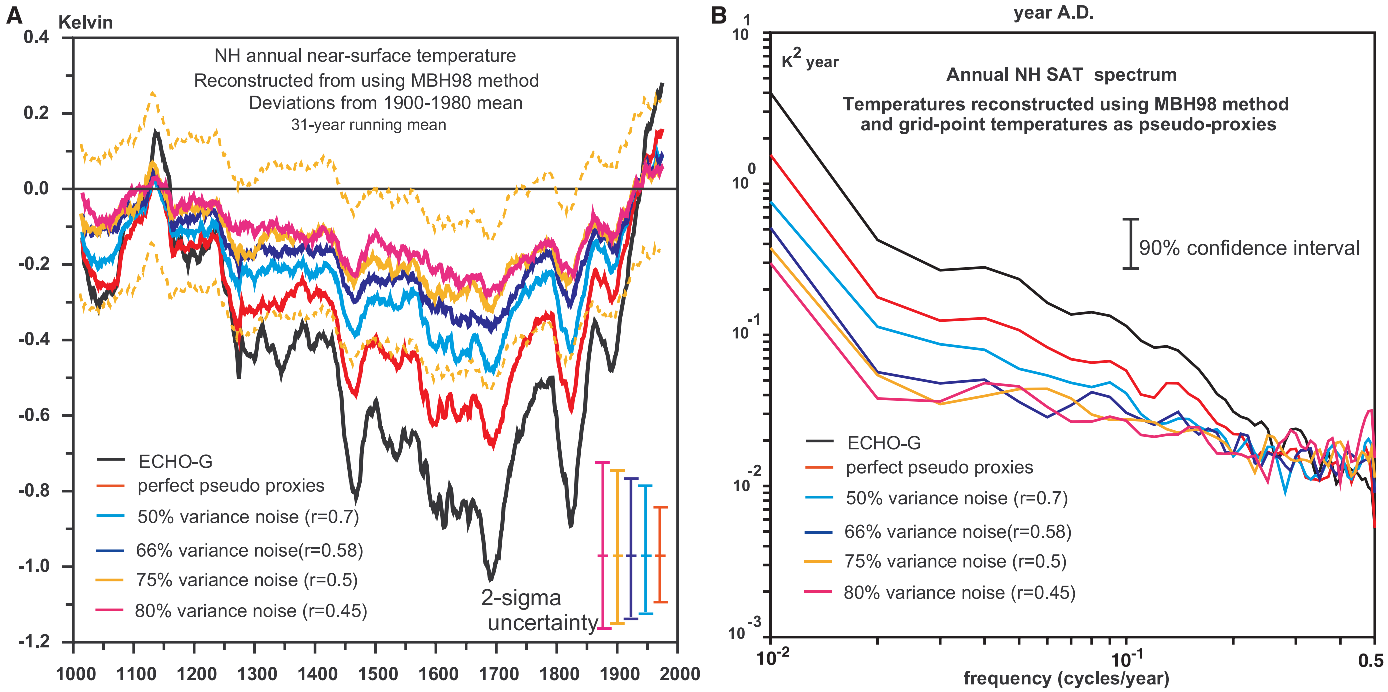
Fig. 2 (**A**) The Northern Hemisphere annual temperature evolution over the last 1000 years. The NH annual temperature simulated by the model ECHO-G and MBH98-reconstructions of this temperature from 105 model grid-points mimicking the multi-proxy network of MBH98. Increasing amounts of noise have been 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 2-sigma uncertainty range (derived as in MBH98 from the variance of the interannual residuals) for the different noise levels is indicated. The reconstruction with ρ=0.5 is shown with its 2×sigma uncertainty range. (**B**) the spectra of the NH annual temperatures shown in (A).

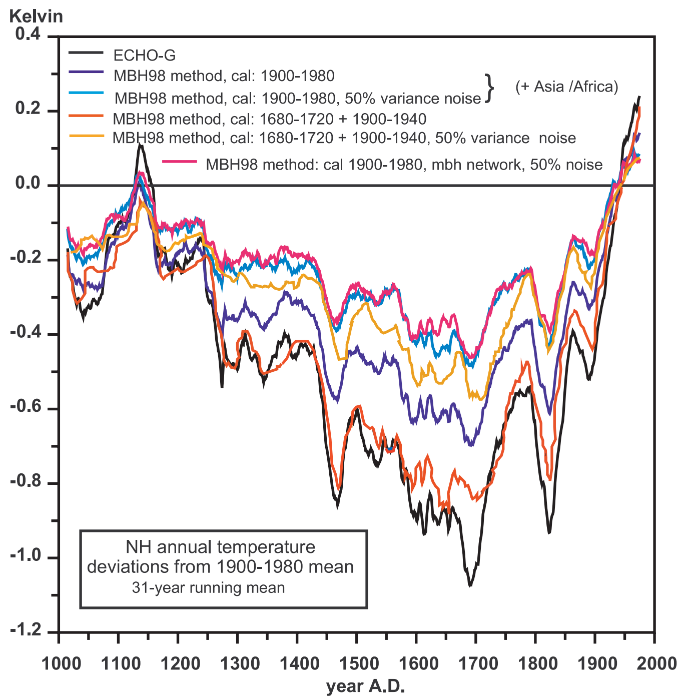
Fig. 3: Simulated (black) and estimated Northern Hemisphere temperature (colors) showing the effect of noise. Different set-ups for the estimation were used: the standard method with 50% noise added (green); with additional pixels in Africa and Asia (locations, Fig. 1) with 50% noise (light blue) and without noise (dark blue); with a different fitting period, namely 1680-1720 plus 1900-1940 (instead of the standard 1900-1980) with 50% noise (light red) and without noise (dark red).

Fig. 4. Simulated NH temperature compared with an estimate with a simple local linear regression on the pseudo-proxies used in Fig. 2. The local temperature is estimated from the each pseudo-proxy and the result is simply averaged over all pseudo-proxies locations to obtain the NH estimate. Also shown is the arithmetic mean of the perfect pseudo-proxies and of the pseudo-proxies containing 50% noise.

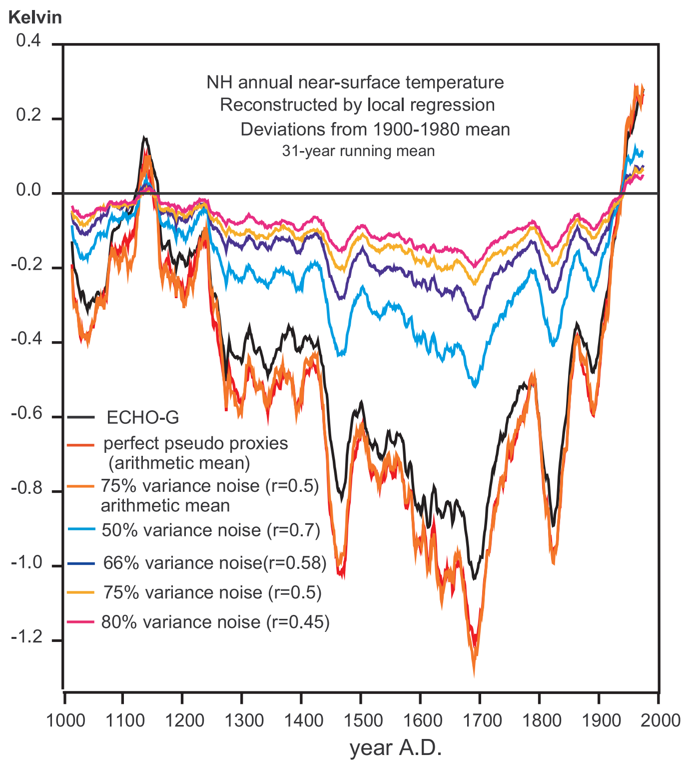


Location of pseudo-proxies in the ECHO-G grid
90
30
-30
-90
-180
-120
-60
0
60
120
180


A
Kelvin
NH annual near-surface temperature
Reconstructed from using MBH98 method
Deviations from 1900-1980 mean
31-year running mean
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
ECHO-G
perfect pseudo proxies
50% variance noise (r=0.7)
66% variance noise(r=0.58)
75% variance noise (r=0.5)
80% variance noise (r=0.45)
2-sigma uncertainty
year A.D.
B
K^2 year
Annual NH SAT spectrum
Temperatures reconstructed using MBH98 method
and grid-point temperatures as pseudo-proxies
90% confidence interval
10^1
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-2}
10^{-1}
0.5
frequency (cycles/year)
ECHO-G
perfect pseudo proxies
50% variance noise (r=0.7)
66% variance noise(r=0.58)
75% variance noise (r=0.5)
80% variance noise (r=0.45)



Kelvin
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
ECHO-G
MBH98 method, cal: 1900-1980
MBH98 method, cal: 1900-1980, 50% variance noise
(+ Asia /Africa)
MBH98 method, cal: 1680-1720 + 1900-1940
MBH98 method, cal: 1680-1720 + 1900-1940, 50% variance noise
MBH98 method: cal 1900-1980, mbh network, 50% noise
NH annual temperature
deviations from 1900-1980 mean
31-year running mean
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
year A.D.


Kelvin
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
NH annual near-surface temperature
Reconstructed by local regression
Deviations from 1900-1980 mean
31-year running mean
ECHO-G
perfect pseudo proxies (arithmetic mean)
75% variance noise (r=0.5) arithmetic mean
50% variance noise (r=0.7)
66% variance noise(r=0.58)
75% variance noise (r=0.5)
80% variance noise (r=0.45)
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
year A.D.