

# Climate reconstruction using ‘Pseudoproxies’

Michael E. Mann and Scott Rutherford

Department of Environmental Sciences, University of Virginia, Charlottesville, VA, USA

Received 11 December 2001; revised 27 March 2002; accepted 2 May 2002; published 31 May 2002.

[1] We test the performance of proxy-based climate field reconstruction methods using sets of synthetic proxy climate indicators. ‘Pseudoproxies’ are constructed through the degradation of instrumental surface temperature data by additive noise with variable statistical properties. Experiments are performed using pseudoproxy networks of varying spatial and seasonal representation and with varying noise attributes. Implications for sampling strategies for improved paleoclimate reconstructions are discussed. *INDEX TERMS*: 1620 Global Change: Climate dynamics (3309); 1694 Global Change: Instruments and techniques; 3309 Meteorology and Atmospheric Dynamics: Climatology (1620); 3344 Meteorology and Atmospheric Dynamics: Paleoclimatology

## 1. Introduction

[2] Reconstructions of global climate in past centuries rely on the ability of relatively sparse sets of proxy data to resolve large-scale patterns of variance in pre-instrumental climate fields [e.g. *Fritts et al.*, 1971; *Bradley and Jones*, 1993; *Overpeck et al.*, 1997; *Mann et al.*, 1998, 1999; *Jones et al.*, 1998]. The skill of such reconstructions can be established through cross-validation with independent withheld instrumental data [e.g., *Mann et al.*, 1998]. Addressing broader questions regarding the sensitivity to spatial sampling, data quality, and prospects for improved reconstructions, however, requires simulation of the paleoclimate reconstruction process.

[3] *Kutzbach and Guetter* [1980] first considered the problem of determining what network of proxy information (precipitation and temperature estimates from pollen) is required to reconstruct large-scale patterns of past climate (sea level pressure). *Bradley* [1996] later examined sampling strategies for proxy-based reconstruction of global mean temperature based on experiments with instrumental and climate model data. *Evans et al.* [1998] considered the impact of sampling distribution and regional signal strength variations in suggesting strategies for reconstructing tropical SST from multiple coral proxy records. To date, however, no studies have systematically simulated the global climate field reconstruction process using synthetic datasets with statistical character emulating that of actual and potentially available global multiproxy networks. Such is the purpose of this study.

## 2. Data and Methods

### 2.1. Instrumental Surface Temperature Data

[4] We make use of all nearly complete (<30% missing annual data, with a requirement of a minimum of 6 months data in a given year) land air/sea surface temperature gridpoint data from 1856–1998 [*Jones et al.*, 1999], with missing data infilled using the Regularized Expectation Maximization (‘RegEm’) technique described by *Schneider* [2000]. This process is used to yield 1312 annual (Jan–Dec) and 1118 warm-season (Apr–Sep) infilled, continuous 143 year gridpoint series. The infilled annual mean surface temperature field will serve as our target for subsequent reconstructions.

### 2.2. Pseudoproxy Data

[5] Networks of pseudoproxy data are generated through the degradation of the instrumental surface temperature gridpoint data by an additive noise component. A given network of  $M_{PRX}$  pseudoproxy indicators of length  $N = 143$  years is constructed from the random selection (without replacement) of  $M_{PRX}$  of the available  $M_{INSTR}$  surface temperature gridpoints. For each of the  $M_{PRX}$  instrumental gridpoint series, we generated an additive noise series as an AR(1) process with autocorrelation as *a priori* specified for that indicator, and with amplitude chosen to insure the signal/noise ratio *a priori* specified for that proxy indicator. The pseudoproxy network is subsequently used in experiments seeking to reconstruct the surface temperature field itself, as described below.

[6] We examine sensitivity to the attributes of the pseudoproxy network in the following ways: 1) We vary the number and location of pseudoproxies to test the sensitivity to spatial sampling, 2) We use a mix of annual and warm-season mean gridpoint series to represent the variable seasonal sensitivity within the proxy network (it is unnecessary to additionally include cold season indicators, since a cold-season half year mean is not statistically independent from a combination of annual and warm half-year information), 3) We vary signal-to-noise (‘SNR’) ratios in the pseudoproxies to test the sensitivity to proxy data quality, 4) We vary the structure of the noise from ‘blue’ to ‘white’ to ‘red’ (through the noise autocorrelation ‘ $\rho$ ’) to test the effects of preferential loss-of-resolved variance by the proxy at high or low frequencies. The noise parameters can either be fixed, or allowed to vary within the network with some specified ensemble variance to simulate the variation in these attributes expected in actual multiproxy data networks. We do not consider the additional possible bias due to age-model uncertainties (i.e., dating errors). It should nonetheless be noted that the indicators used in previous multiproxy reconstructions by *Mann et al.* [1998, 1999] were based almost exclusively on multiply replicated chronologies, minimizing the influence of this latter potential source of bias.

### 2.3. Calibration and Cross-Validation

[7] Annual mean temperature patterns are reconstructed based on calibration of the (potentially seasonally-mixed) information in the pseudoproxy network against the  $M = 1312$  annual mean surface temperature gridpoint series during a training period of duration equal to half the full available (143 year) interval. A split calibration/verification procedure was used wherein both the 1856–1927 and 1928–1998 periods were alternatively used for both calibration and for verification (i.e., independent testing of the reconstructions with withheld instrumental data). The results obtained were similar in both cases. We focus, however, on the ‘early miss’ experiments in which the later data are used for calibration, and for which the influence of non-stationarity in the climate is minimized [see *Rutherford et al.*, 2002].

[8] The calibration technique is a generalization of the *Mann et al.* [1998, 1999] approach making more complete use of the data covariance information [see *Schneider*, 2001; *Rutherford et al.*, 2002]. Calibration is accomplished through estimating the correlation matrix of the merged pseudoproxy + instrumental surface temperature data series. The reconstruction uses the calibrated covariance information to estimate the (missing) gridpoint temper-

**Table 1.** Reconstruction Experiments for Different Pseudoproxy Network Indicating Experiment Group (w/ Number of Experiments in Parentheses), Number of Pseudoproxies ( $M_{PRX}$ ), % of Surface Temperature Gridpoint Data Represented, % Annual (vs Warm-Season) Indicators in Network (A%), Ensemble Mean Signal-to-Noise Ratio (SNR), Ensemble Mean Noise Autocorrelation ( $\rho$ ), and RE and CE Statistics for Full Field (' $Mlt$ ') and Global Mean (' $Glb$ ') Series

#	$M_{PR}$	EARLY		VERIF		RE		CE	
		%	A%	SNR	$\rho$	$Mlt$	$Glb$	$Mlt$	$Glb$
1 (2)	656	50	100	$\infty$	0.0	0.480	0.970	0.270	0.870
2 (2)				0.5	0.0	0.340	0.945	0.090	0.640
3 (2)	367	70	100	$\infty$	0.0	0.470	0.970	0.260	0.800
4 (2)				0.5	0.5	0.330	0.935	0.065	0.570
5 (2)	197	85	100	$\infty$	0.0	0.445	0.985	0.225	0.880
6 (4)				0.5		0.280	0.925	0.00	0.525
7 (3)			50	0.5		0.275	0.930	-0.01	0.520
8 (2)					0.5	0.290	0.910	0.01	0.375
9 (2)					-0.5	0.305	0.890	0.055	0.470
10(2)	112	91.5	100	$\infty$	0.0	0.415	0.960	0.185	0.765
11(1)			30			0.385	0.965	0.145	0.780
12(2)				0.7		0.305	0.925	0.045	0.610
13(4)				0.5		0.275	0.875	0.015	0.335
14(3)			50			0.270	0.845	0.010	0.305
15(2)			30			0.265	0.845	0.015	0.370
16(2)			100		0.5	0.280	0.860	0.015	0.290
17(2)					-0.5	0.280	0.855	0.030	0.380
18(3)				0.4	0	0.250	0.810	-0.01	0.240
18A						0.240	0.780	-0.02	0.125
18B						0.245	0.840	-0.01	0.375
19(2)			50		0.5	0.255	0.810	-0.005	0.185
19A						0.255	0.790	0.005	0.235
20(2)			100		0.9	0.300	0.855	0.035	0.210
21(2)				0.3		0.200	0.690	-0.030	0.080
22(2)	63	95.0	100	$\infty$	0	0.380	0.955	0.14	0.72
23(2)				0.5		0.245	0.785	0.00	0.18
24(2)	37	98.5	100	$\infty$	0	0.360	0.920	0.12	0.75
26(1)	20	95.5	100	$\infty$	0	0.320	0.860	0.095	0.62
28(1)	13	99	100	$\infty$	0	0.275	0.750	0.055	0.185
29(1)			50	0.5		0.12	0.305	0.00	0.04

ature from only the (available) pseudoproxy indicators during the independent reconstruction (and, in this case, verification) interval.

[9] For the purposes of simplicity, only temperature indicators are considered. Actual multiproxy data are potentially sensitive to other climate variables (e.g., precipitation). Since the calibration method makes no assumptions of a local relationship between the proxy indicator and the large-scale temperature field, the impact of non-temperature influences is largely accommodated through SNRs which vary within the pseudoproxy indicator ensemble. For example, large precipitation anomalies are typically associated with a local low pressure anomaly, which is in turn associated, albeit not necessarily in a simple linear manner, with significant advective surface temperature variations in other regions. Thus, a precipitation-sensitive proxy indicator provides information about the large-scale temperature field, though with a non-local and potentially weaker effective signal.

[10] We made use of two distinct diagnostics of the data variance resolved by the reconstructions [see e.g. *Cook et al.*, 1994] including the reduction of error ('RE') which uses the calibration period climatology as a baseline (this is called ' $\beta$ ' by *Mann et al.*, [1999]), and the coefficient of efficiency ('CE') which instead employs the verification period climatology as a baseline:

$$RE = 1 - \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (\hat{y}_i - \bar{y}_{ic})^2} \quad (1)$$

$$CE = 1 - \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (\hat{y}_i - \bar{y}_{iv})^2} \quad (2)$$

$RE = 0$  and  $CE = 0$  represent the scores for a reconstruction which simply specifies the climatological mean, and thus only positive values of the statistics indicate statistically significant skill.  $CE = 0$  is a more challenging threshold since, unlike  $RE$ ,  $CE$  does not

reward reconstructing an observed change in mean relative to the calibration period. Verification  $RE$  and  $CE$  scores are computed for both global domain mean temperature (' $Glb$ ') and for the full multivariate field [' $Mlt$ ']—in this latter case the sums in (1) and (2) extend both over time and over the  $M_{INST}$  annual-mean temperature gridpoints].

[11] Uncertainties in the reconstructions can be estimated from the residual uncalibrated verification period variance. We also examine the spectra of calibration residuals for evidence of systematic bias associated with preferential loss of reconstructed variance at high or low frequencies.

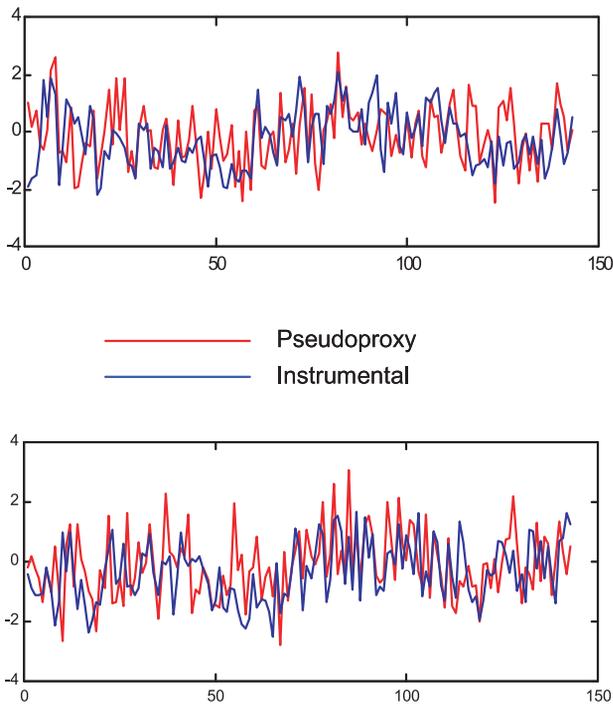
### 3. Results

[12] A number of experiments were performed with pseudoproxy networks of varying sampling density, SNR, noise autocorrelation, and mix of seasonal influence. Table 1 summarizes results from a number of these experiments in which these sampling/noise parameters were varied. Allowing the noise parameters to vary randomly in space within each experiment yielded results indistinguishable from experiments in which they were fixed at an ensemble mean so that, for simplicity, results from the former are not shown. Some general observations from these experiments are:

i) Reconstructive skill is insensitive to the mix of seasonal vs. annual information over a fairly large fractional range.

ii) Reconstructive skill is largely insensitive to the spectrum of the noise, though, as discussed below, the spectrum of verification residuals (which has implications for the nature of uncertainties in reconstructed values) is not.

iii) At denser concentrations of available pseudoproxy networks, there is relatively little sensitivity to the precise distribution of the pseudoproxies. At lower concentrations (e.g., less than 100



**Figure 1.** Sample ‘Pseudoproxy’ series for two gridpoint locations (top) 37.5N, 22.5E and (bottom) 12.5N, 32.5W from experiment 18A (Table 1) along with the associated actual instrumental annual-mean temperature series.

pseudoproxies) this is less true. There is more discussion of this in the context of a specific example below.

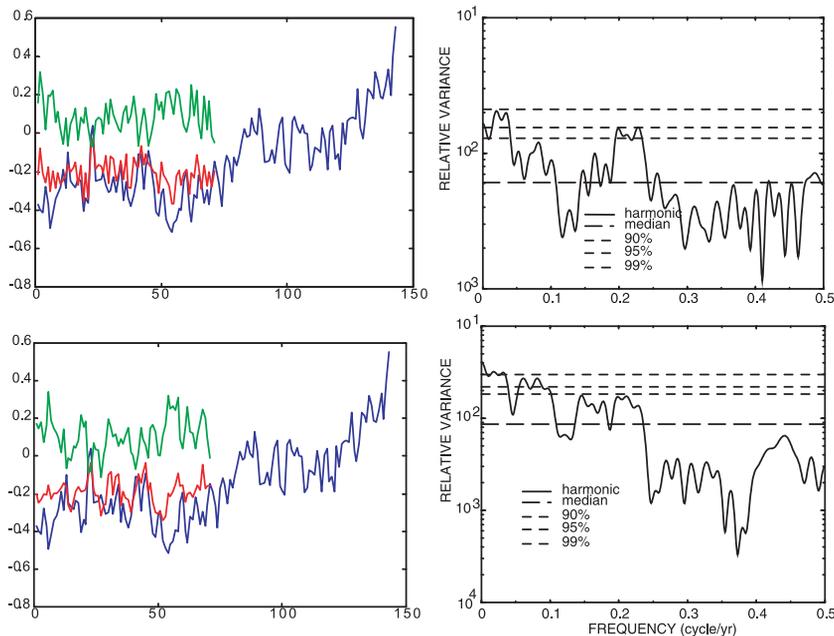
[13] In Figure 1 we show two sample ‘pseudoproxy’ indicators along with the corresponding instrumental annual gridpoint surface temperature series taken from one such experiment (18A: white noise component w/ SNR = 0.4, 112 annual-mean-only temperature pseudoproxies available). Visually, the comparison between a typical pseudoproxy and corresponding surface temperature series

closely resembles that for actual proxy data. Indeed, an SNR of 0.4 is equal to an average shared variance between proxy and instrumental gridpoint of  $r = 0.29$ , ( $r^2 = 0.08$ ), which is typical of estimates of local temperature correlations with multiproxy indicators [see e.g. Jones *et al.*, 1998].

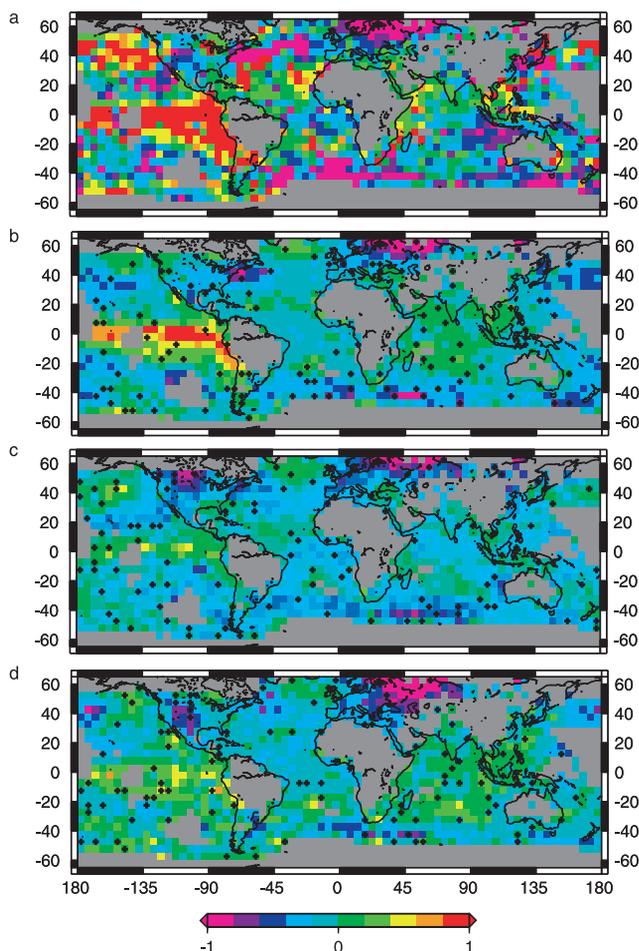
[14] Global mean temperature reconstructions are compared against the actual infilled verification period global mean temperature series (Figure 2). Also shown are the corresponding residuals and their associated spectra. For pseudoproxies with a white noise component, the spectrum of residuals is itself approximately white. For pseudoproxies with a red noise component, on the other hand, the spectrum of residuals shows similar evidence of redness. If actual proxy indicators have a red noise error component, which is equivalent to the preferential loss of lower-frequency information, estimates of uncertainties in the longest timescales of variation will be larger than is indicated from the nominal unresolved variance, and must be taken into account in estimating appropriate confidence intervals [e.g. Mann *et al.*, 1999].

[15] At relatively modest percentages of missing data, the reconstructive skill is relatively insensitive to the precise spatial sampling (because most spatial degrees of freedom in the field are sampled). However, greater sensitivity is observed at lower sampling densities (particularly at lower SNR. The resolvability of ENSO-scale temperature patterns of variance depend somewhat sensitively, for example, on how well the eastern tropical Pacific is sampled. Figure 3 compares the actual infilled instrumental surface temperature pattern for 1877(a), a known strong El Nino year, along w/ reconstructions based on 112 randomly distributed pseudoproxies with (b) no noise, and with (c, d) SNR of 0.4 for two different random realizations. Note that distribution (c), with greater sampling in the eastern and central tropical Pacific, more faithfully captures the El Nino-related pattern of warmth for that year than another distribution (b) with the same global sampling density and SNR.

[16] It is of interest to compare the results of the pseudoproxy experiments with those obtained for actual multiproxy reconstructions [Mann *et al.*, 1998, 1999]. For example, experiment 19A in Table 1, with SNL = 0.4,  $\rho = 0.5$ , and 50/50 split between annual vs. warm-season indicators compares favorably with the results obtained using the ‘RegEm’ approach on e.g. the same 112 indicators proxy indicators of Mann *et al.* [1998] that are available



**Figure 2.** Global mean surface temperature reconstruction (red) along with actual infilled data (blue) for experiments 18A and 19A. Residuals are shown (green) along with the corresponding spectrum of residuals (right) based on multiple-taper procedure described by Mann and Lees [1996].



**Figure 3.** Actual (a) and Reconstructed Surface Temperature Anomaly Pattern (relative to 1961–1990 mean) for the year 1877 From Experiments 10B (b), 18A (c) and 18B (d), illustrating the influence of both SNR and the specific locations of pseudoproxies on the fidelity of the reconstructed pattern.

back to 1820 ( $Glb = 0.82$ ,  $Mlt = 0.25$ ). Though the intent here is to address the underlying sampling issues as generally as possible, it is worth noting that an experiment using the approximate actual locations of the 112 indicators used by Mann *et al.* [1998] yields quite similar results ( $Glb = 0.85$ ,  $Mlt = 0.26$ ). Taking this as a baseline estimate of the fidelity of global surface temperature reconstructions possible with currently available global multiproxy networks, it is possible to use the results of these experiments (i.e. Table 1) to gauge prospects for improved reconstructions based either on the acquisition of more widespread or higher quality proxy data. For example, similar levels of increased skill ( $Mlt RE = 0.3$ ) might be expected, roughly speaking, given either (i) the same number of proxy indicators (112) as used by Mann *et al.* [1998], with the average SNR were increased from 0.4 to 0.7 (i.e., experiment 12) or (ii) the SNR kept approximately the same (0.5) but the number of indicators nearly doubled to 197 (e.g. experiments 7–9). Of course, at this level of sparseness, as shown earlier, the particular set of locations of the proxy data can be a quite significant factor. Moreover, in actual paleoclimate reconstruction, there are additional practical constraints governing the possible distribution of such locations (e.g., where coral records are actually available).

#### 4. Conclusions

[17] The results of the experiments using synthetic ‘pseudo-proxy’ indicators derived from the instrumental record itself to

reconstruct instrumental surface temperature patterns over independent intervals in time support conclusions from cross-validation results for actual proxy-reconstructions. These results furthermore place upper limits on the possible resolved variance in proxy-based paleoclimate reconstruction, based on experiments using ‘perfect’ (noise-free) proxies. Finally, the results of these experiments support at least two distinct strategies for improved proxy-based large-scale surface temperature pattern reconstructions. Similar levels of resolved spatial variance can be obtained by a strategy of widespread sampling (hundreds of proxies) with relatively low SNRs, or more selective sampling with a smaller number of proxy indicators with higher signal-to-noise ratios. In the latter case, in particular, certain regions are likely to be of particular importance (e.g. the tropical Pacific). Initial experiments with forced and control model simulations using noise-free indicators are described by Rutherford *et al.* [2002]. Similar experiments generating long synthetic proxy climate time series are currently underway. Such experiments should allow us to investigate implications on longer timescales that cannot be addressed with the experiments, such as are described here, based on resampling of the instrumental record.

[18] **Acknowledgments.** This work was supported by the NOAA and NSF-sponsored Earth Systems History program.

#### References

- Bradley, R. S., Are there optimum sites for global paleotemperature reconstruction?, in *NATO ASI Series Climatic Variations and Forcing Mechanisms of the last 2000 Years*, edited by P. D. Jones, R. S. Bradley, and J. Jouzel, pp. 603–624, Springer-Verlag, Berlin Heidelberg, 1996.
- Bradley, R. S., and P. D. Jones, ‘Little Ice Age’ summer temperature variations: their nature and relevance to recent global warming trends, *The Holocene*, 3, 367–376, 1993.
- Cook, E. R., K. R. Briffa, and P. D. Jones, Spatial Regression Methods in *Dendroclimatology: A Review and Comparison of Two Techniques*, *International Journal of Climatology*, 14, 379–402, 1994.
- Evans, M. N., A. Kaplan, and M. A. Cane, Optimal Sites for Coral-based Reconstruction of Global Sea Surface Temperature, *Paleoceanography*, 13, 502–516, 1998.
- Fritts, H. C., T. J. Blasing, B. P. Hayden, and J. E. Kutzbach, Multivariate techniques for specifying tree-growth and climate relationships and for reconstructing anomalies in paleoclimate, *Journal of Applied Meteorology*, 10, 845–864, 1971.
- Jones, P. D., K. R. Briffa, T. P. Barnett, and S. F. B. Tett, High-Resolution Paleoclimatic Records for the Last Millennium: Interpretation, Integration and Comparison with Circulation Model Control-Run Temperatures, *The Holocene*, 8, 455–471, 1998.
- Jones, P. D., M. New, D. E. Parker, S. Martin, and J. G. Rigor, Surface Air Temperature and its Changes over the Past 150 Years, *Reviews of Geophysics*, 37, 173–199, 1999.
- Kutzbach, J. E., and P. J. Guetter, On the design of paleoenvironmental data networks for estimating large-scale patterns of climate, *Quaternary Research*, 14, 169–187, 1980.
- Mann, M. E., and J. M. Lees, Robust Estimation of Background Noise and Signal Detection in Climatic Time Series, *Climate Change*, 33, 409–445, 1996.
- Mann, M. E., R. S. Bradley, and M. K. Hughes, Global-scale temperature patterns and climate forcing over the past six centuries, *Nature*, 392, 779–787, 1998.
- Mann, M. E., R. S. Bradley, and M. K. Hughes, Northern Hemisphere Temperatures During the Past Millennium: Inferences, Uncertainties, and Limitations, *Geophysical Research Letters*, 26, 759–762, 1999.
- Overpeck, J., et al., Arctic Environmental Change of the Last Four Centuries, *Science*, 278, 1251–1256, 1997.
- Rutherford, S., M. E. Mann, T. L. Delworth, and R. Stouffer, The Performance of Covariance-Based Methods of Climate Field Reconstruction Under Stationary and Nonstationary Forcing, *J. Climate*, in revision, 2002.
- Schneider, T., Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values, *Journal of Climate*, 14, 853–887, 2001.

M. E. Mann and S. Rutherford, Department of Environmental Sciences, University of Virginia, Charlottesville, VA 22902, USA. (mann@virginia.edu)