Borehole climate reconstructions: Spatial structure and hemispheric averages

Henry N. Pollack and Jason E. Smerdon
Department of Geological Sciences, University of Michigan, Ann Arbor, Michigan, USA

Received 18 September 2003; revised 10 February 2004; accepted 3 March 2004; published 5 June 2004.

[1] Ground surface temperature (GST) reconstructions determined from temperature profiles measured in terrestrial boreholes, when averaged over the Northern Hemisphere, estimate a surface warming of ~1 K during the interval AD 1500–2000. Other traditional proxy-based estimates suggest less warming during the same interval. *Mann et al.* [2003a] have raised two issues with regard to borehole-based reconstructions. The first focuses on the need for spatial gridding and area-weighting of the ensemble of borehole-based GST reconstructions to yield an average hemispheric reconstruction. The second asserts that application of optimal detection techniques show that the GST only weakly displays the spatial structure of the surface air temperature (SAT). We demonstrate the consistency of GST warming estimates by showing that over a wide range of grid element area and occupancy weighting schemes, the five-century GST change falls in the range of 0.89–1.05 K. We examine the subhemispheric spatial correlation of GST and SAT trends at various spatial scales. In the 5-degree grid employed for optimal detection, we find that the majority of grid element means are determined from three or fewer boreholes, a number that is insufficient to suppress site-specific noise via ensemble averaging. Significant spatial correlation between SAT and GST emerges in a 5-degree grid if low-occupancy grid elements are excluded, and also in a 30-degree grid in which grid element means are better determined through higher occupancy. Reconstructions assembled after excluding low-occupancy grid elements show a five-century GST change in the range of 1.02–1.06 K. *INDEX TERMS:* 1645 Global Change: Solid Earth; 1699 Global Change: General or miscellaneous; 3309 Meteorology and Atmospheric Dynamics: Climatology (1620); 3322 Meteorology and Atmospheric Dynamics: Land/atmosphere interactions; 3344 Meteorology and Atmospheric Dynamics: Paleoclimatology; *KEYWORDS:* boreholes, paleoclimate, spatial analysis, surface temperature

Citation: Pollack, H. N., and J. E. Smerdon (2004), Borehole climate reconstructions: Spatial structure and hemispheric averages, *J. Geophys. Res.*, *109*, D11106, doi:10.1029/2003JD004163.

1. Introduction

[2] Paleoclimate reconstructions that describe global and hemispheric surface temperature histories over the past two millennia include several inferred from traditional surface temperature proxies [*Overpeck et al.*, 1997; *Jones et al.*, 1998; *Mann et al.*, 1999; *Briffa et al.*, 2001; *Crowley and Lowery*, 2000; *Esper et al.*, 2002; *Mann and Jones*, 2003]. Other investigators have used direct measurements of subsurface temperatures in boreholes to infer ground surface temperature (GST) histories [*Huang et al.*, 2000; *Harris and Chapman*, 2001; *Beltrami*, 2002a]. *Briffa and Osborn* [2002] discuss several reconstructions for the Northern Hemisphere, all of which show the 20th century as a period of warming of at least 0.5 K, but in the 400 years preceding the 20th century the reconstructions show more variability.

[3] Many reasons have been suggested for the differences in the various reconstructions: different proxies respond to different target seasons; the calibration of some proxies may imperfectly separate temperature from other climatic factors that affect the proxy; the spatial distributions of data are different in the various reconstructions; different proxies have better fidelity at different temporal frequencies; ground and air temperatures may respond differently to land-use or climatic changes; and local data have been spatially aggregated to form a hemispheric average by different methods. Cogent discussions of these issues have been presented by *Beltrami* [2002b], *Briffa et al.* [2001], *Harris and Chapman* [2001], *Briffa and Osborn* [2002], *Esper et al.* [2002], *Trenberth and Otto-Bliesner* [2003], *Mann et al.* [2003b] and *Zorita et al.* [2003]. In this paper we first address the aggregation issue, particularly as it pertains to the regional and hemispheric climate histories estimated from individual

0148-0227/04/2003JD004163$09.00

borehole temperature profiles. Subsequently we address the question of sub-hemispheric spatial structure in the hemispheric ensemble of borehole reconstructions.

2. Aggregating Reconstructions

[4] It is important to note initially that most of the following discussion is not about analysis of subsurface temperature profiles, the primary data that borehole investigators use to reconstruct a surface temperature history. Rather it addresses the spatial aggregation and representation of the ensemble of reconstructions that have been derived in one particular analysis of borehole temperatures, i.e., the analysis by *Huang et al.* [2000]. Other analyses of the primary temperature data to reconstruct a global or hemispheric GST history have been presented by *Harris and Chapman* [2001] and *Beltrami* [2002a].

[5] Borehole investigators have aggregated individual borehole reconstructions in different ways. *Pollack et al.* [1998] and *Huang et al.* [2000] estimated global GST histories as simple arithmetic averages of 358 and 616 individual borehole site reconstructions, respectively. The five-century Northern Hemisphere reconstruction (453 sites) of *Huang et al.* [2000] has subsequently been reassessed by *Briffa and Osborn* [2002], who placed the borehole sites into a latitude-longitude grid and estimated a hemispheric GST history from the grid element means, but without any weighting by grid element area or occupancy (T. Osborn, personal communication, 2003). The *Huang et al.* [2000] hemispheric reconstruction has also been reassessed by *Mann et al.* [2003a, hereinafter referred to as M03], who calculated, inter alia, an area-weighted hemispheric GST history (see *Rutherford and Mann* [2004] for a correction).

[6] *Harris and Chapman* [2001], in their independent analysis of the borehole temperature profiles, placed a 5-degree grid over the Northern Hemisphere borehole data sites, and within each grid element averaged the transient temperature profiles. In the subsequent analysis of the grid element averages to establish a Northern Hemisphere reconstruction, the grid element averages were area-weighted (D. Chapman, personal communication, 2003). The fact that the Harris and Chapman gridded hemispheric analysis yielded a temperature change very similar to that determined by *Huang et al.* [2000] through simple averaging of individual site reconstructions, offers a first hint that the differences existing between borehole and traditional proxy-based reconstructions probably cannot be attributed simply to the different ways in which they have been aggregated.

3. Hemispheric Reconstructions Weighted by Grid Element Area

[7] To determine the range of hemispheric reconstructions that result from different aggregation schemes, we compute five-century GST reconstructions for the Northern Hemisphere as a simple arithmetic mean of the individual site reconstructions, and as an area-weighted mean of grid element averages for all grid elements containing boreholes. We use the current contents of the publicly available database of borehole temperatures and climate reconstructions [*Huang and Pollack*, 1998] that presently includes 695 sites in the Northern Hemisphere, as displayed in Figure 1. This database

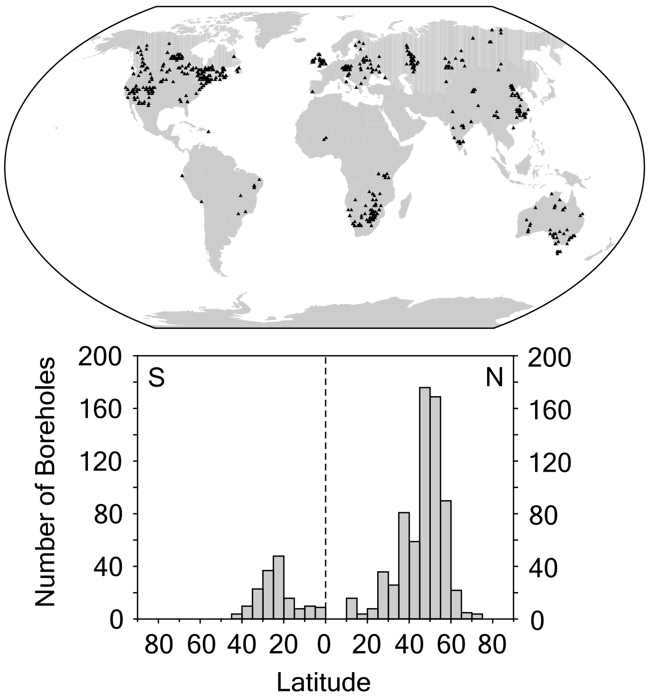

Figure 1. Locations of boreholes in the global database, with histogram of latitudinal distribution (modified after *Harris and Chapman* [2001] to reflect the current status of the database).

includes both raw borehole temperature data, as well as a derived GST reconstruction comprising linear temperature trends for each of the past five centuries, at each borehole location.

[8] We have examined grids of various sizes: 5, 10, 15, 30, 45, and 90 degrees, with latitude incremented from the equator and longitude from the Greenwich meridian. Grid size plays a variable role in assembling a hemispheric reconstruction. A very small grid size ($\ll$1 degree) is tantamount to enclosing each data site in its own grid element. A hemispheric reconstruction from such a small grid is, in effect, a weighted average of the individual site reconstructions, with the weightings being proportional to the cosine of the site latitude. At the other extreme, a single "grid element" comprising the entire hemisphere would contain all data sites in a single element, thereby rendering area weighting moot, and yielding a hemispheric mean that is the simple average of all the data sites in the hemisphere. The hemispheric reconstructions of *Huang et al.* [2000] are effectively the latter.

[9] The results of our area-weighting experiments are shown in Table 1 and Figure 2 (see also Appendix A for details of our area-weighting calculations). The simple averaging of the 695 individual site reconstructions over the hemisphere yields 1.04 K of warming over the past five centuries, whereas the gridded and area-weighted reconstructions range from 0.89 to 1.05 K, depending on the grid size. The simple arithmetic average reconstruction, without any area weighting, lies very near the cooler bound of the reconstruction envelope. The previously mentioned *Harris and Chapman* [2001] investigation, a hybrid analysis based on both borehole observations and the SAT historical

Table 1. Five-Century Area-Weighted GST Changes in the Northern Hemisphere

Grid Size, deg	695 Sites $\Delta T_{1500-2000}$, K	453 Sites $\Delta T_{1500-2000}$, K
5	1.02	0.96
10	1.02	0.95
15	0.94	0.86
30	0.89	0.83
45	0.93	0.94
90	1.05	1.09

record, also obtained a gridded and area-weighted reconstruction showing circa 1 K of warming over the past five centuries.

[10] For direct comparison with the analyses of *Briffa and Osborn* [2002] and M03 we performed the same calculations with the smaller data set comprising only the 453 Northern Hemisphere sites used by those investigators and by *Huang et al.* [2000]. These calculations (also shown in Table 1) show that the differences arising from the size of the data set are inconsequential: simple averaging of the 453 individual site reconstructions yields a five-century warming over the hemisphere of 1.06 K, and gridded and area-weighted averages in the range 0.83–1.09 K, results not significantly different from the 695 sites we analyze here. We were not able to reproduce the area-weighted reconstruction shown in M03, using the same data set distributed in a 5-degree grid, or any other grid size. *Rutherford and Mann* [2004] acknowledge that there was an error in the M03 calculation of the area-weighted reconstruction.

4. Hemispheric Reconstructions Weighted by Area and Grid Element Population

[11] The discussion of aggregation should not rest with the observations made above, because area-weighting alone does not take into account the statistical robustness of each grid element mean. Each borehole yields a site-specific ground surface temperature reconstruction using a subsurface temperature profile that contains both a climatologic signal and site-specific nonclimatologic noise. A grid element mean reconstruction estimated from many individual site reconstructions is likely to be better determined than a "mean" estimated from a single site, because of the cancellation of site-specific nonclimatologic effects that accompanies the averaging of individual site reconstructions. Discussions of the sources of such noise have been presented by *Shen et al.* [1995] and *Pollack and Huang* [2000].

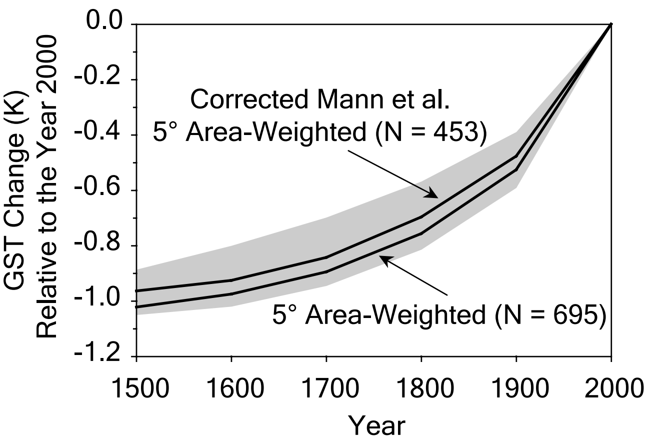

Figure 2. Five-century area-weighted reconstructions of ground surface temperature history from 695 boreholes in the Northern Hemisphere. Shaded area represents the range occupied by six hemispheric reconstructions in grids ranging in size from 5 × 5 to 90 × 90 degrees. The corrected 5-degree area-weighted reconstruction of *Mann et al.* [2003a] [*Rutherford and Mann*, 2004], and our 5-degree area-weighted reconstruction using a larger data set are superimposed.

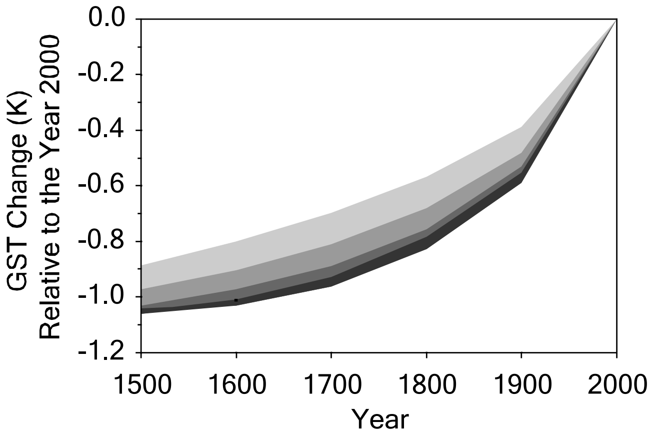

Figure 3. Five-century area- and occupancy-weighted reconstructions of ground surface temperature history from 695 boreholes in the Northern Hemisphere. The entire shaded region represents the range of area-weighted hemispheric reconstructions in grids of various sizes, as shown in Figure 2. The sequentially darker regions represent progressively heavier weighting by occupancy (see text). Each darker region overlies part of the lighter shading beneath.

[12] In forming a hemispheric or global reconstruction, additional weighting beyond simple area weighting should be applied to reflect the variable robustness of the estimate of each grid element mean. In addition to weighting by the area of the grid element discussed above, we also weight by grid element occupancy level. We apply an a empirical weighting of the form $A \cdot N^w$, where A is the area of the grid element, N is the number of sites in the grid element, and w is an exponent running through the values 0, 0.25, 0.5, 0.75, and 1. The weighting with w = 0 corresponds to a uniform occupancy weighting of all grid element means, independent of the number of sites in each grid element. It is equivalent to the simple area-weighted reconstruction discussed above and shown in Figure 2. The fractional values of w between 0 and 1 give increasing weight to well-populated grid elements; the weighting w = 1 gives each grid element a weight equal to its occupancy level. This empirical weighting scheme spans a wide range of weightings by both area and occupancy.

[13] Reconstructions weighted both by occupancy and grid element area are displayed in Figure 3. These results are rather insensitive to the size of the grid and to longitudinal translation of the grid over the hemisphere. Progres-

sively increased weighting by grid element population leads to a convergence of reconstructions very close to that represented by the simple arithmetic mean of the individual site reconstructions. It is clear that grid element means, weighted either by area alone or by both area and population, lead to the conclusion that borehole reconstructions indicate circa 1 K of warming in the Northern Hemisphere over the past five centuries. The corrected area-weighted estimate of warming by *Rutherford and Mann* [2004] is fully consistent with this conclusion.

5. Subhemispheric Spatial Structure

[14] M03 have also argued that the ensemble of GST reconstructions does not show very much of the subhemispheric spatial structure displayed in the 20th century SAT record, and that much of what has been identified in the geothermal observations as climatic warming, both in the 20th century and in the previous four centuries, is probably noise and/or bias. M03 acknowledge that for the 20th century, the simple arithmetic mean of borehole site reconstructions over the Northern Hemisphere presented by *Huang et al.* [2000] compares favorably to the mean trend of the SAT data. They go on to point out, however, that a comparison of colocated 5-degree grid element means shows prominent differences between the two data sets, with more than a quarter of the grid elements even showing discrepant signs, i.e., one data set shows a warming trend for a grid element in which the other data set indicates cooling. Their subsequent empirical orthogonal function (EOF) analysis of GST reconstructions detects only a weak SAT signal. In the following discussion we argue that the M03 treatment of the geothermal reconstructions is tenuous, and therefore their conclusions stemming from that analysis must be viewed with caution.

[15] Initially we compare the spatial characteristics of the GST and SAT on a 5-degree grid, the only grid size investigated by M03. Our analysis includes all 695 borehole sites presently available in the Northern Hemisphere data set, a number more than 50% greater than the number of sites analyzed by M03. The comparison focuses on those 5-degree grid elements that contain both borehole reconstructions and 20th century SAT data. In each grid element we calculate the mean temperature change of both GST and SAT over the 20th century, using the borehole reconstructions of GST and the University of East Anglia compilation of SAT data [*Jones et al.*, 1999]. Figure 4a displays the colocated mean temperature changes in each grid element, determined from GST and SAT. The two data sets display similar variability, principally in the range of −1 to +2 K/century, although most points are situated in the quadrant in which both the SAT and GST show 20th century warming. The correlation between the SAT and GST estimates in the larger data set is 0.13, significant at the p = 0.09 level (all significance levels we cite are one-sided), not very different than the 0.11 correlation determined by M03 (no significance level given) using the smaller data set. These numbers indicate only a weak correlation, with significance outside the 95% confidence interval.

[16] If in fact our analysis ended at this point, we would conclude, as did M03, that there is only weak spatial similarity between the two arrays. It is not difficult, however, to determine the source of the disconnect: in a 5-degree grid, the grid element means of the borehole reconstructions are very weakly determined because there are only a small number of occupants in most of the grid elements. Even with the larger data set we analyze here (695 sites versus 453 sites analyzed by M03), almost 1 in 5 of the occupied 5-degree grid elements contain only a single borehole, and fully half of the grid elements contain three or fewer boreholes. As a simple example of the

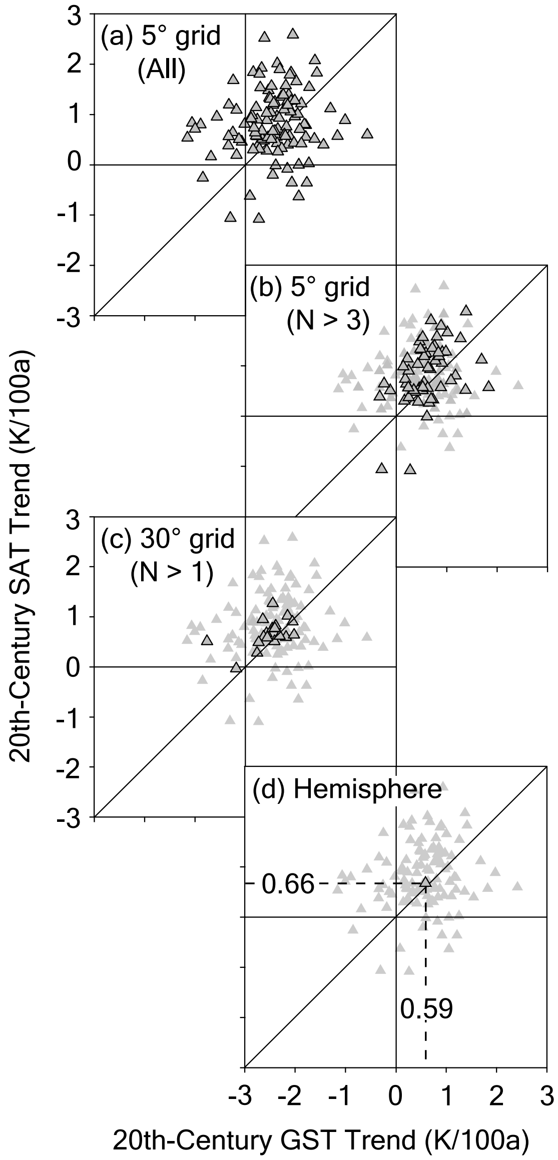

Figure 4. Comparison of 20th century Northern Hemisphere temperature trends in grid elements of different sizes and selected populations, as determined from SAT and borehole reconstructions in (a) all occupied 5-degree grid elements, (b) 5-degree grid elements with more than three occupants, (c) 30-degree grid elements containing more than one occupant, and (d) complete Northern Hemisphere. The 5-degree pattern in Figure 4a is shown faintly in the background of Figures 4b, 4c, and 4d to enable easy visual comparisons.

consequences of low occupancy, we note that of the 25 grid elements displaying discrepant signs between GST and SAT trends, 20 contain three or fewer boreholes.

5.1. Signal Enhancement in Borehole Analysis

[17] To appreciate the significance of this sparse occupancy, one must recognize the uncertainties that affect the interpretation of individual borehole temperature profiles. The inevitable presence of site-specific noise in subsurface temperatures (see *Shen et al.* [1995] and *Pollack and Huang* [2000] for a discussion of these effects), has led geothermal researchers to caution against placing confidence in a reconstruction derived from a single borehole; more typically they discuss regional ensembles of data [*Shen et al.*, 1995; *Pollack et al.*, 1996] and the regional climatic reconstructions derived from the data ensembles. This procedure has parallels in the assembly of a dendroclimatological chronology. A chronology derived from a single tree would surely be viewed with skepticism compared to a chronology assembled from scores of trees.

[18] To enhance the regional climate signal while suppressing site-specific noise, geothermal researchers employ simultaneous inversion of several boreholes [*Beltrami et al.*, 1997] or averaging of several individual reconstructions [*Shen et al.*, 1995]. Simultaneous inversion seeks a signal compatible (within tolerance levels) with all of the individual borehole temperature profiles; the averaging technique operates under the assumption that site-specific noise is random and cancels under averaging, whereas a regional climate signal is common and additive. Under many circumstances the two procedures, simultaneous inversion and ensemble averaging, yield very similar results [*Pollack et al.*, 1996].

[19] At the 5-degree spatial scale, the individual GST reconstructions are only thinly aggregated, and accordingly the grid element means of the borehole site reconstructions do not benefit from the noise suppression that accompanies aggregation. Accordingly, many of the grid element means are burdened with large uncertainties.

5.2. Spatial Correlation Enhancement

[20] The problems associated with sparse occupancy can be addressed, and enhancement of the spatial correlation between the 20th century GST and SAT can be achieved, in at least two ways, both of which we examine here. The first retains for analysis only those 5-degree grid element means that meet some occupancy threshold, and excludes the highly uncertain grid element means estimated from very few occupants (often only a single occupant). The second method of enhancement adopts a larger grid size so that each grid element has the potential to encompass a greater number of boreholes, thereby avoiding the situation where the available borehole sites are distributed thinly into many grid elements. Both methods enable a search for common spatial structure between the GST and SAT, the first at the 5-degree scale but with a data set truncated by low-occupancy considerations, and the second at a larger spatial scale.

5.2.1. Truncation of Low-Occupancy Grid Elements in a 5-Degree Grid

[21] The 5-degree grid divides the hemisphere into 1296 grid elements, 113 of which are occupied by boreholes and also have SAT records. Of the occupied grid elements, 21 contain only a single borehole, 21 contain two boreholes, and 16 contain three. Together these low-occupancy grid elements represent more than half of the occupied elements but contain only 111 boreholes out of the available 695, i.e., 51% of the grid elements represent only 16% of the data set. This makes clear the potential for a gravely distorted correlation analysis if the low-occupancy grid elements are used. Removing the low-occupancy grid elements from the analysis allows 84% of the data set to be retained in 55 grid elements.

Table 2. GST-SAT Correlation Statistics for Selected Data Sets

	Correlation, r	Significance[a]	Mean RMS Error, K/100a
	695 Data Set		
5° grid (all)	0.13	0.09	0.71
5° grid ($N > 3$)	0.42	0.0007	0.53
5° grid ($N < 4$)	0.016	0.45	0.88
30° grid (all)	0.45	0.03	0.31
30° grid ($N > 1$)	0.50	0.017	0.29
	453 Data Set		
5° grid M03 (all)	0.11	not given	1.0
5° grid M03 "optimal" (all)	0.25	0.01	not given

[a]One-sided p value.

[22] We compare both the sparsely populated and the more densely populated subsets of the 5-degree gridded borehole data to the corresponding SAT (see Table 2). The sparsely populated subset, when compared to the SAT, shows an RMS difference of 0.88 K/century, and a very low correlation of 0.016 with poor statistical significance ($p = 0.45$). The more densely populated subset (Figure 4b), in which 55 grid elements contain more than three boreholes, shows an RMS difference of 0.53 K/century, and a correlation of 0.42, significant at the $p = 0.0007$ level. Following the M03 analysis, we report all significance statistics using the maximum number of degrees of freedom determined from the number of grid elements represented in each correlation. These results clearly show that low-occupancy grid elements obscure the significant correlation that exists between GST and SAT. As we have noted before, the underlying reason for this obfuscation is that grid element means estimated from small numbers of occupants are generally weak. Those grid elements with greater occupancy and better determined means correlate better with the SAT.

[23] The "optimal detection" EOF analysis of M03 also effects a separation of signal from noise and/or bias. In their EOF analysis, the first eigenvector has been characterized as representing noise and/or bias, and the second eigenvector as carrying the climate signal contained in the borehole data. The correlation and significance levels of their "optimally detected" climate signal and our "$N > 3$" signal are 0.25 ($p = 0.01$) and 0.42 ($p = 0.0007$), respectively. The differences arise largely from the manner in which noise is addressed in the respective analyses. M03 analyzed the borehole data set only on the 5-degree grid, retained all of the data, and employed a relative error-weighting scheme based on occupancy. However, because there are so many 5-degree grid elements that are sparsely occupied, they play a disproportionate and degrading role in detecting the "optimal" signal. To illustrate this point

ad absurdum one can imagine a data distribution such that every grid element contained only a single borehole, each determining the "mean" for that grid element. Occupancy weighting in this case would be uniform, and yet say nothing about the relative accuracy of the grid element means. A relative weighting scheme such as that employed by M03 will in this case fail to recognize the poor quality of the grid element means; it will simply weight them equally. Yet a correlation analysis of weakly determined grid element means is unlikely to yield anything but a weak result.

[24] In contrast, we have simply excluded low-occupancy grid elements from the analysis, and seek a climatologic signal in the higher-occupancy grid elements. In essence this exclusion gives zero weight to low-occupancy grid elements. The culling by occupancy need not be severe; in the 5-degree grid we exclude only those grid elements with three or fewer occupants. As we have shown, the remaining higher-occupancy grid elements yield a statistically significant correlation with the SAT.

5.2.2. Larger Grid Size

[25] Enhanced spatial correlation can also be demonstrated at larger grid sizes with accordingly higher occupancy levels in many of the grid elements. For example, a 30-degree grid divides the hemisphere into 36 grid elements, 19 of which are occupied by borehole sites. Only one grid element contains a single borehole; the median number of occupants in the occupied grid elements is 17, as compared to only 3 in the 5-degree grid.

[26] The correlation between SAT and GST with no grid elements excluded improves from r = 0.13 on a 5-degree grid to 0.45 on the 30-degree grid, and with an improved significance ($p = 0.03$; see Table 2). The mean RMS misfit is reduced from 0.71 K/century to 0.31 K/century between the 5-degree and 30-degree grids, a reduction of 56%. Clearly, the greater the number of boreholes in a grid element, the better the GST compares to the SAT. Conversely, the fewer the number of occupants the less robust is the mean, and the poorer the comparison with the SAT.

[27] When we exclude the single-occupancy grid element in the 30-degree grid, the comparison of GST to SAT continues to improve (see Table 2 and Figure 4c). The two variables show an RMS difference of 0.29 K/century and a correlation of 0.50, significant at the $p = 0.017$ level. The contraction of the array of colocated GST and SAT data seen in Figure 4c illustrates the reduction of the misfit between the estimated GST and SAT trends.

[28] Considering the full hemisphere as a single grid element (Figure 4d) yields the previously known and well-recognized result that there is good correspondence between GST and SAT 20th century trends (0.59 and 0.66 K/century, respectively) at the hemispheric scale. On a 5-degree grid in which grid elements with three or fewer occupants are excluded, the GST hemispheric estimate increases to 0.61 K/century and the agreement is even closer.

[29] It is well known that the SAT exhibits spatial correlation, with a correlation scale length on an annual basis in the range 1000–2000 km [*Hansen and Lebedeff*, 1987; *Briffa and Jones*, 1993; *Mann and Park*, 1994; *Jones et al.*, 1997]. Moreover, SAT correlation scale lengths are timescale-dependent [*Jones et al.*, 1997], with the century scale length greater than the annual scale length. These considerations suggest a reduction in the number of spatial degrees of freedom that will influence the statistical significance of a correlation between the GST and SAT. The correlations we report between GST and SAT remain significant within the 90% confidence interval for many fewer degrees of freedom. The 0.42 correlation between GST and SAT in the 5-degree grid is significant down to approximately 9 spatial degrees of freedom and the 0.50 correlation between GST and SAT in the 30-degree grid is significant down to approximately 7 spatial degrees of freedom.

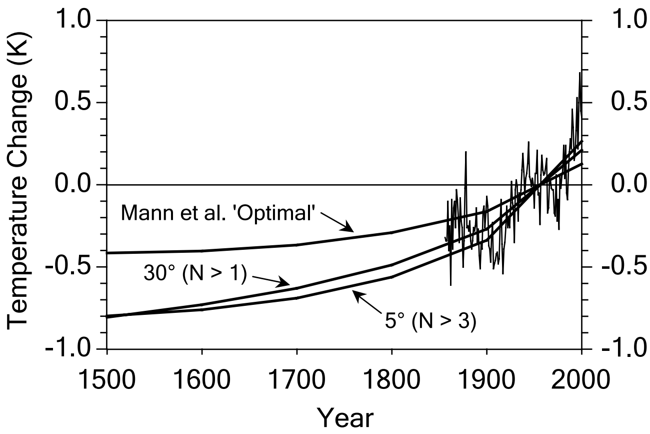

Figure 5. Five- and thirty-degree borehole reconstructions for the Northern Hemisphere (low-occupancy grid elements excluded). Additional occupancy weighting of the retained grid elements leads to a convergence of the two reconstructions. Also shown are the corrected *Mann et al.* [2003a] "optimal" reconstruction [*Rutherford and Mann*, 2004], and the 20th century SAT annual anomaly series [*Jones et al.*, 1999]. The SAT is referenced to its mean in the interval 1961–1990; the other plots are referenced to the zero crossing of the SAT 20th century trend.

[30] Figure 5 displays two area-weighted five-century GST reconstructions corresponding to the occupancy-filtered 5-degree and 30-degree data sets, both well correlated with the SAT in the 20th century. These two reconstructions are very similar, suggesting that both approaches to dealing with the problems presented by low-occupancy are effective. When additional occupancy weighting is applied to the retained grid elements, the two reconstructions converge, and lie well within the uncertainty envelope of the simple arithmetic mean of all 695 borehole sites. Also shown in Figure 5 is the corrected "optimal" reconstruction of *Mann et al.* [2003a] [*Rutherford and Mann*, 2004]. There remains a significant unexplained difference between our reconstructions and the corrected "optimal" reconstruction of Mann et al.

[31] We conclude from this discussion that given the current size of the borehole database and the nature of the associated noise, attempts to detect similar spatial patterns in the SAT and GST will face difficulties when many grid element means are weakly determined because of low grid element occupancy, i.e., small sample size. We have presented two ways to avoid this problem: (1) discarding low-occupancy grid elements from a 5-degree grid before

correlations are determined between the GST and SAT; and (2) employing a larger grid that yields higher occupancy by distributing the data sites into a smaller number of larger grid elements.

6. Possibilities of Bias?

[32] M03, in addition to arguing that borehole reconstructions are burdened with noise, also suggest that they may be systematically biased. In general, the borehole reconstruction of *Huang et al.* [2000] and the various reconstructions and aggregations discussed above show that the 16th through 18th centuries were, on average, significantly colder than depicted in some of the traditional proxy-based reconstructions. Therefore the bias M03 intimate would be some characteristic in the borehole data or analysis that has led to an over-estimation of climatic warming since the year 1500. We examine several possibilities below.

6.1. Borehole Logging Dates

[33] The borehole temperature inversions of *Huang et al.* [2000] were parameterized to estimate century-long trends of surface temperature [*Huang et al.*, 1996], even though the median year of temperature logging in the borehole database is 1983, some two decades prior to the end of the century. Recognizing the range of logging dates, *Harris and Chapman* [2001] forward continued all of the subsurface transient profiles to a common date near the end of the century, by assuming a constant surface temperature from the time of logging. Both the *Huang et al.* [2000] and *Harris and Chapman* [2001] reconstructions must be considered as muted estimates of the 20th century trend, *Huang et al.* [2000] because many borehole sites at the time of measurement would not yet have been exposed to the significant warming of the last two decades of the 20th century, and Harris and Chapman because their extrapolation scenario is an underestimate of the actual development of the surface air temperature. Nevertheless, this end-of-century bias present to some degree in both analyses, if adjusted for, would yield an even greater difference between the borehole and proxy-based reconstructions, rather than bring them closer together.

6.2. Borehole Site Geography

[34] That the borehole sites comprise a predominantly midlatitude continental data set is clearly shown in Figure 1. Bias arising from this limited geographic sampling is of course possible. If the five-century hemispheric warming were actually less than what the midlatitude boreholes suggest, then regions outside the midlatitude continents would have to exhibit less warming than indicated in the midlatitudes. A plethora of evidence including the SAT indicates 20th century warming in high latitudes is similar to or even greater than that exhibited in the midlatitudes. Thus sampling of high-latitude geography is unlikely to reduce the estimate of hemispheric warming derived from a largely midlatitude data set. A determination that low-latitude regions or the oceans have experienced a more moderate warming history than the middle- and high-latitude continents, so as to yield a full hemispheric result that differs from a geographically limited continental reconstruction, is of course yet another possible avenue of reconciliation.

[35] The *Esper et al.* [2002] and *Briffa et al.* [2001] reconstructions have been characterized as midlatitude continental reconstructions also. These two reconstructions describe temperature histories that are in closest agreement with borehole-based estimates. The similarity of these three estimates, given their similar geographical settings, lends some additional confidence in the viability of each. It also suggests that the dendrochronology construction techniques that Esper et al. and Briffa et al. developed to preserve low-frequency trends in their climate reconstructions may well have been necessary. We also note, however, that these reconstructions may be biased due to volcanism [*Shindell et al.*, 2003].

6.3. Analysis of Different Data Sets

[36] We have searched for a possible bias arising from our use of a larger data set than that analyzed by M03. We analyzed the smaller set used by M03, using the same gridding, weighting and aggregation experiments reported here for the larger set of borehole reconstructions (see Table 1 and Figures 2 and 3). Aside from different error estimates due to the smaller sample size, our results from the smaller data set are essentially the same as the results from the larger in all major characteristics. The acquisition of new data from different geographic localities has not significantly altered the estimates of temperature change derived from the smaller ensemble of reconstructions.

6.4. Seasonal Filters

[37] It is well observed that mean annual ground surface temperature differs from the mean annual SAT in regions where there is snow cover and/or seasonal freezing and thawing, due to the insulation effects of snow cover and the latent heat released or absorbed [*Gosnold et al.*, 1997; *Hinkel et al.*, 2001; *Smerdon et al.*, 2003a]. It is also clear that the coupling of SAT and GST over a single year is complex: summer evapotranspiration cools the ground whereas winter snow cover and soil freezing partially decouples the ground from winter air temperatures [*Kane et al.*, 2001; *Sokratov and Barry*, 2002; *Lin et al.*, 2003; *Smerdon et al.*, 2003b; *Mann and Schmidt*, 2003; *Chapman et al.*, 2004]. In high-latitude regions where the mean annual temperature is below freezing, the latent heat of fusion prevents the deeper subsurface from seeing the warmth of summer [*Hinkel et al.*, 2001]. M. G. Bartlett et al. (Snow and the ground temperature record of climate change, submitted to *Journal of Geophysical Research*, 2004) show that the interannual variability of snow cover onset, thickness and duration can lead to a highly variable decoupling, sometimes warming the ground and sometimes cooling the ground relative to the SAT. *Stieglitz et al.* [2003] note the effect of decadal variations in snow cover on subsurface temperatures in the Alaskan Arctic, but even stable decadal variability, which penetrates to no greater than 50–60 m depth, probably has little influence on the estimation of multicentury trends that are imprinted on borehole temperature profiles to depths of several

hundred meters. These studies collectively illustrate that it is not easy to make a case for systematic bias in subsurface temperatures due to these processes.

[38] The more relevant discussion is not whether the GST and SAT are identical (it is clear that they are not) but rather, do they show similar trends over long time periods. Long-term simulations of both air and soil temperatures from global three-dimensional coupled climate models [*González-Rouco et al.*, 2003] are indicating that seasonal differences in coupling do exist, but are of little significance over long timescales. *González-Rouco et al.* [2003] comment that the deep soil temperature is a good proxy for the annual SAT on continents, and variations in the two temperatures are almost indistinguishable. Additionally, the simulations by *González-Rouco et al.* [2003] suggest a hemispheric SAT history that is very similar to the GST history recovered from terrestrial boreholes.

6.5. Investigator Bias

[39] Another possible source of bias might be investigator bias. It is abstractly conceivable that some aspect of the processing and/or interpretation of the borehole temperature observations could impart a bias. The analysis of the borehole data requires a separation of the steady state component of the subsurface temperature field that is associated with the outward flux of heat from Earth's interior. The residual temperatures after this separation must then be interpreted in terms of a climatic signal and nonclimatic noise. The interpretive methodology (e.g., the method of inversion) therefore might be examined for bias. Nevertheless, the possibility of investigator bias affecting the five-century reconstruction of the surface temperature is highly unlikely. This assertion is supported by the fact that the subsurface temperature observations in the public database have been analyzed independently by three separate geothermal research groups, each with their own and rather different interpretive methods (*Huang et al.* [2000]; *Harris and Chapman* [2001]; *Beltrami* [2002a]; the M03 critique was based specifically on the results of the Huang et al. analysis). These three independent estimates of the five-century change in GST are very similar. The likelihood of parallel biases in each of these analyses therefore seems small.

7. Conclusions

[40] The principal outcome of the analyses we have presented is that the five-century reconstruction of the ground surface temperature from boreholes located in the Northern Hemisphere shows warming of about 1 K, and is essentially independent of the aggregation and weighting scheme employed to determine the hemispheric mean. Significant spatial correlation between GST reconstructions and 20th century SAT trends emerges after careful pruning of low occupancy grid elements, or by gridding at a larger scale so that grid elements generally have higher occupancy. Accordingly, the hemispheric reconstruction derived from borehole temperatures continues to stand as an independent and robust estimate of changes in continental temperatures over the past five centuries.

Appendix A

[41] We calculate the area of each grid element on a unit sphere as

$$A(\lambda_1, \lambda_2, \theta_1, \theta_2) = \int\int dA = \int_{\lambda_1}^{\lambda_2}\int_{\theta_1}^{\theta_2} \cos(\lambda) d\lambda d\theta$$
$$= [\sin(\lambda_2) - \sin(\lambda_1)]\Delta\theta,$$

where λ and θ are latitude and longitude, respectively, λ_1, λ_2, θ_1 and θ_2 are the bounding latitudes and longitudes of the grid element, and $\Delta\lambda$, $\Delta\theta$ are the differences between the bounding latitudes and bounding longitudes, respectively.

[42] The equivalence of the above equation to the "cosine-latitude" weighting referred to in the work of *Mann et al.* [2003a] is easily demonstrated by employing the trigonometric identity

$$\sin(\lambda_2) - \sin(\lambda_1) = 2\cos\left(\frac{\lambda_1 + \lambda_2}{2}\right)\sin\left(\frac{\lambda_2 - \lambda_1}{2}\right),$$

which leads directly to

$$A(\lambda_1, \lambda_2, \theta_1, \theta_2) = 2\cos(\overline{\lambda})\sin(\Delta\lambda/2)\Delta\theta,$$

where $\overline{\lambda}$ is $(\lambda_1 + \lambda_2)/2$. Thus the area of a grid element is seen to be proportional to $\cos(\overline{\lambda})$.

[43] We calculate the area-weighted hemispheric mean (for each century or for the sum of all five centuries) as

$$\frac{\sum_{i=1}^{N} f(\lambda_i, \theta_i) \cdot w(\lambda_i, \theta_i)}{\sum_{i=1}^{N} w(\lambda_i, \theta_i)},$$

where $f(\lambda_i, \theta_i)$ is the ith grid element mean, $w(\lambda_i, \theta_i)$ is the weighting (area, occupancy or both) of the ith grid element, and the summations are over the N occupied grid elements. This expression is the well-recognized definition of a weighted average.

[44] **Acknowledgments.** We thank David Chapman, Rob Harris, and Shaopeng Huang for comments on an early draft of this paper and four reviewers for their helpful critiques. This research was supported in part by NSF grant ATM-0081864 and NASA grant GEWEC-0000-0132 and by the Vice President for Research at the University of Michigan.

References

Beltrami, H. (2002a), Climate from borehole data: Energy fluxes and temperatures since 1500, *Geophys. Res. Lett.*, *29*(23), 2111, doi:10.1029/2002GL015702.

Beltrami, H. (2002b), Earth's long-term memory, *Science*, *297*, 206–207.

Beltrami, H., L. Z. Cheng, and J.-C. Mareschal (1997), Simultaneous inversion of borehole temperature data for past climate determination, *Geophys. J. Int.*, *129*, 311–318.

Briffa, K. R., and P. D. Jones (1993), Surface air temperature variations during the 20th century: Part 2—Implications for large-scale high-frequency paleoclimate studies, *Holocene*, *3*, 82–92.

Briffa, K. R., and T. J. Osborn (2002), Blowing hot and cold, *Science*, *295*, 2227–2228.

Briffa, K. R., T. J. Osborn, F. H. Schweingruber, I. C. Harris, P. D. Jones, S. G. Shiyatov, and E. A. Vaganov (2001), Low-frequency temperature variations from a northern tree-ring density network, *J. Geophys. Res.*, *106*, 2929–2941.

Chapman, D. S., M. G. Bartlett, and R. N. Harris (2004), Comment on "Ground vs. surface air temperature trends: Implications for borehole surface temperature reconstructions" by M. E. Mann and G. Schmidt, *Geophys. Res. Lett.*, *31*, L07205, doi:10.1029/2003GL019054.

Crowley, T. J., and T. S. Lowery (2000), How warm was the Medieval Warm Period?, *Ambio.*, *29*, 51–54.
Esper, J., E. R. Cook, and F. H. Schweingruber (2002), Low-frequency signals in long tree-line chronologies for reconstructing past temperature variability, *Science*, *295*, 2250–2253.
González-Rouco, F., H. von Storch, and E. Zorita (2003), Deep soil temperature as proxy for surface air-temperature in a coupled model simulation of the last thousand years, *Geophys. Res. Lett.*, *30*(21), 2116, doi:10.1029/2003GL018264.
Gosnold, W. D., P. Todhunter, and W. Schmidt (1997), The borehole temperature record of climate warming in the mid-continent of North America, *Global Planet. Change*, *15*, 33–45.
Hansen, J., and S. Lebedeff (1987), Global trends of measured surface air temperature, *J. Geophys. Res.*, *92*, 13,345–13,372.
Harris, R. N., and D. S. Chapman (2001), Mid-latitude (30°–60°N) climatic warming inferred by combining borehole temperatures with surface air temperatures, *Geophys. Res. Lett.*, *28*, 747–750.
Hinkel, K. M., F. Paetzold, F. E. Nelson, and J. G. Bockheim (2001), Patterns of soil temperature and moisture in the active layer and upper permafrost at Barrow, Alaska: 1993–1999, *Global Planet. Change*, *29*, 293–309.
Huang, S., and H. N. Pollack (1998), Global borehole temperature database for climate reconstruction, *Contrib. Ser. 1998-044*, IGBP Pages/World Data Cent. A for Paleoclimatol. Data, NOAA/NGDC Paleoclimatol. Program, Boulder, Colo.
Huang, S., P.-Y. Shen, and H. N. Pollack (1996), Deriving century-long trends of surface temperature from borehole temperatures, *Geophys. Res. Lett.*, *23*, 257–260.
Huang, S., H. N. Pollack, and P.-Y. Shen (2000), Temperature trends over the last five centuries reconstructed from borehole temperatures, *Nature*, *403*, 756–758.
Jones, P. D., T. J. Osborn, and K. R. Briffa (1997), Estimating sampling errors in large-scale temperature averages, *J. Clim.*, *10*, 2548–2568.
Jones, P. D., K. R. Briffa, T. P. Barnett, and S. F. B. Tett (1998), High-resolution paleoclimatic records for the last millennium: Integration, interpretation, and comparison with general circulation model control run temperatures, *Holocene*, *8*, 455–471.
Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Rigor (1999), Surface air temperature and its variations over the last 150 years, *Rev. Geophys.*, *37*, 173–199.
Kane, D. L., K. M. Hinkel, D. J. Goering, L. D. Hinzman, and S. I. Outcalt (2001), Non-conductive heat transfer associated with frozen soils, *Global Planet. Change*, *29*, 275–292.
Lin, X., J. E. Smerdon, A. W. England, and H. N. Pollack (2003), A model study of the effects of climatic precipitation changes on ground temperatures, *J. Geophys. Res.*, *108*(D7), 4230, doi:10.1029/2002JD002878.
Mann, M. E., and P. D. Jones (2003), Global surface temperatures over the past two millennia, *Geophys. Res. Lett.*, *30*(15), 1820, doi:10.1029/2003GL017814.
Mann, M. E., and J. Park (1994), Global-scale modes of surface temperature variability on interannual to century timescales, *J. Geophys. Res.*, *99*, 25,819–25,833.
Mann, M. E., and G. Schmidt (2003), Ground vs. surface air temperature trends: Implications for borehole surface temperature reconstructions, *Geophys. Res. Lett.*, *30*(12), 1607, doi:10.1029/2003GL017170.
Mann, M. E., R. S. Bradley, and M. K. Hughes (1999), Northern Hemisphere temperatures during the last millennium: Inferences, uncertainties, and limitations, *Geophys. Res. Lett.*, *26*, 759–762.
Mann, M. E., S. Rutherford, R. S. Bradley, M. K. Hughes, and F. T. Keimig (2003a), Optimal surface temperature reconstructions using terrestrial borehole data, *J. Geophys. Res.*, *108*(D7), 4203, doi:10.1029/2002JD002532.
Mann, M. E., et al. (2003b), On past temperatures and anomalous late-20th century warmth, *Eos Trans. AGU*, *84*, 256–258.
Overpeck, J., et al. (1997), Arctic environmental change of the last four centuries, *Science*, *278*, 1251–1256.
Pollack, H. N., and S. Huang (2000), Climate reconstruction from subsurface temperatures, *Annu. Rev. Earth Planet. Sci.*, *28*, 339–365.
Pollack, H. N., P.-Y. Shen, and S. Huang (1996), Inference of ground surface temperature history from subsurface temperature data: Interpreting ensembles of borehole logs, *Pure Appl. Geophys.*, *147*, 537–550.
Pollack, H. N., S. Huang, and P.-Y. Shen (1998), Climate change record in subsurface temperatures: A global perspective, *Science*, *282*, 279–281.
Rutherford, S., and M. E. Mann (2004), Correction to "Optimal surface temperature reconstructions using terrestrial borehole data," *J. Geophys. Res.*, *109*, D11107, doi:10.1029/2003JD004290.
Shen, P.-Y., H. N. Pollack, S. Huang, and K. Wang (1995), Effects of subsurface heterogeneity on the inference of climate change from borehole temperature data: Models studies and field examples from Canada, *J. Geophys. Res.*, *100*, 6383–6396.
Shindell, D. T., G. A. Schmidt, R. L. Miller, and M. E. Mann (2003), Volcanic and solar forcing of climate change during the preindustrial era, *J. Clim.*, *16*, 4094–4107.
Smerdon, J. E., H. N. Pollack, J. W. Enz, and M. J. Lewis (2003a), Conduction-dominated heat transport of the annual temperature signal in soil, *J. Geophys. Res.*, *108*(B9), 2431, doi:10.1029/2002JB002351.
Smerdon, J. E., H. N. Pollack, M. Kresl, and J. F. Wehmiller (2003b), Conductive transport of annual temperature signals in variable subsurface media and regional analyses of seasonal decoupling between air and ground surface temperatures, *Eos Trans. AGU*, *84*(46), Fall Meet. Suppl., F926.
Sokratov, S. A., and R. G. Barry (2002), Intraseasonal variation in the thermoinsulation effect of snow cover on soil temperatures and energy balance, *J. Geophys. Res.*, *107*(D10), 4093, doi:10.1029/2001JD000489.
Stieglitz, M., S. J. Dery, V. E. Romanovsky, and T. E. Osterkamp (2003), The role of snow cover in the warming of arctic permafrost, *Geophys. Res. Lett.*, *30*(13), 1721, doi:10.1029/2003GL017337.
Trenberth, K. E., and B. L. Otto-Bliesner (2003), Toward integrated reconstruction of past climates, *Science*, *300*, 589–591.
Zorita, E., F. Gonzalez-Rouco, and S. Legutke (2003), Testing the Mann et al. (1998) approach to paleoclimate reconstructions in the context of a 1000-yr control simulation with the ECHO-G coupled climate model, *J. Clim.*, *16*, 1378–1390.

H. N. Pollack and J. E. Smerdon, Department of Geological Sciences, University of Michigan, Ann Arbor, MI 48109-1063, USA. (hpollack@umich.edu)