The Holocene 6,3 (1996) pp. 255–263

Estimates of low frequency natural variability in near-surface air temperature

T.P. Barnett,[1] B.D. Santer,[2] P.D. Jones,[3] R.S. Bradley[4] and K.R. Briffa[3]

([1]Scripps Institution of Oceanography, UCSD, La Jolla, CA 92093, USA; [2]Lawrence Livermore National Laboratory, Livermore, CA 94550, USA; [3]University of East Anglia, Norwich NR4 7TJ, UK; [4]University of Massachusetts, Amherst, MA 01003, USA)

Received 17 July 1995; revised manuscript accepted 12 January 1996

A HOLOCENE RESEARCH PAPER

Abstract: Estimates of the spectrum of natural variability are critical to the problem of detecting an anthropogenic signal in global climate observations. Without such information it is impossible to say that current climate change is different or unique from changes that have happened in the past and, therefore, potentially due to man-induced causes. We have estimated the spectrum of natural variability from a globally distributed set of palaeo-temperature proxies and compared it with comparable estimates from two long control integrations of coupled general circulation models – the type used to predict anthropogenic change due to greenhouse gases. None of the three estimates of the natural variability spectrum agree with each other on the low-frequency, near-global time/space scales. Until this dichotomy is resolved, it will be hard to say, with confidence, that an anthropogenic climate signal has or has not been detected.

Key words: Air temperature, natural variability, anthropogenic effects, climatic change, general circulation models, GCM, global climate.

Introduction

Recent studies have attempted to detect a human-induced effect on global climate, relying on numerical model estimates of natural climate variability as their 'yardsticks' for judging the significance of observed changes in climate (Stouffer *et al.*, 1994; Santer *et al.*, 1996b; Hegerl *et al.*, 1996; Manabe and Stouffer, 1996; see also Wigley and Barnett, 1990 and Santer *et al.*, 1996c, for a recent summary of the detection problem). Indeed, it can be argued that *the* key to detecting an anthropogenic signal is quantitative knowledge of the spectrum of natural variability. Without such information, it is impossible to say if an observed climate change is different or unique from changes that have happened in the past and, therefore, potentially due to anthropogenic causes.

The purpose of this paper is to investigate quantitatively how well the model estimates of natural variability at timescales of several decades-to-centuries and planetary scales agree with those obtained from palaeo data. Prior to making such comparisons, it is important to summarize the potential shortcomings of both types of noise estimators, as well as the problems associated with using instrumental observations for the same purpose.

Ideally, one could use instrumental data to estimate the spectrum of natural variability. Unfortunately, the instrumental record is barely a century long for the best-observed of variables (e.g. near-surface temperatures). Further, it is almost certain that the component of 'total' natural variability in this record (i.e. internally generated variability, plus that due to changes in solar output, volcanic aerosols etc.) is convolved with some human-induced signal (e.g. trace gases plus anthropogenic aerosols). There seems to be no valid empirical way to remove this anthropogenic effect. Attempts to remove such an effect theoretically and/or with numerical model results requires faith that the underlying model is itself correct; a considerable assumption for many models. For these reasons, we cannot directly use the instrumental record to estimate natural variability and, hence, need to turn to indirect methods.

Model estimates

The natural variability, or 'noise', used in the above-referenced investigations is the variability simulated by Coupled Atmosphere-Ocean General Circulation Model (CGCM) control runs with fixed levels of greenhouse gases. We refer to this as 'model' noise, to distinguish the estimates from 'observed' noise. With increased computer resources, such control runs are now of up to 1000 years in length, thus permitting estimates of internally generated natural variability on timescales of decades to centuries. Attention has been devoted to the question of how well current CGCMs simulate interannual to decadal timescale natural variability (Stouffer *et al.*, 1994; Latif and Barnett, 1994; 1996; Mehta and Delworth, 1996; Delworth *et al.*, 1993; Manabe and Stouffer, 1996). In general, the models appear to perform well in regions of the North Pacific and North Atlantic, i.e. spatial scales considerably smaller than the greenhouse signal. However, little work has been done on evaluating the reliability of model noise on multidecadal time and planetary spatial scales, the crucial scales for detection. It is well known that the CGCM control runs omit known aspects of variability important on the interannual to century timescales, so it is highly likely that the estimate of internally generated natural variability on these timescales will be low. For instance, control integrations do not incorporate changes in solar output or changes in volcanically induced aerosol inputs to the atmosphere. In fact, it is not clear how the former could ever be incorporated, given our limited knowledge of past solar forcing histories. Thus, the variability that the models capture is due solely to the complex interactions between the atmosphere, with its very fast timescales and the ocean and ice with their longer timescales. Even these sources of variability in the CGCMs are deficient, since they do not simulate ENSO events adequately (too weak). So they reproduce only weakly one of the largest known principal modes of interannual ocean/atmosphere variability. Why this is so is presently unknown. However, for present purposes, the 3–5-year ENSO timescale is short compared to the decadal to century timescales we shall be investigating, and so this omission may not be critical for present purposes (although the relatively high frequency forcing could conceivably generate a lower frequency response in analogy to the ideas of Hasselmann, 1976).

In summary, there are known problems with most current CGCMs. It is exactly for that reason that we want to see how well they simulate the low-frequency noise that must be known for the detection problem.

Palaeo noise estimates

It seems reasonable to verify the model estimates of noise with long (> 300–500 years) palaeoclimatic reconstructions that encompass temperature fluctuations prior to any substantial human influence on climate. Unfortunately, palaeo data are not perfect for this purpose as we shall see below.

Measured tree-ring-width or density series from individual trees, ice cores and/or coral temperature proxies frequently exhibit relatively long timescale area-related trends. These trends are thought to represent properties of the palaeo proxy itself, e.g. the biological growth function in individual trees. Because such trends do not represent a response to climate change, they must be removed before the data are useful. Otherwise, they would bias the subsequent climate reconstructions. Removing these trends, however, also removes potential climate information on relatively long timescales. The extent to which this occurs depends on the nature of the original data and the chronology construction methods employed. For instance, some of the tree records will better reflect multidecadal to century timescale variability than others (see Briffa *et al.*, 1992a).

Even when 'calibrating' proxy data by regression against instrumental records, only part (approximately at most 50–60%; Briffa and Jones, 1993) of the observed variance is captured. This means palaeo data almost certainly underestimate the magnitude of the spectrum of natural variability. The real hope for validating the models with palaeo data thus rests in comparing *patterns* of spatial variabiilty and the temporal spectra.

A further problem is 'scale incompatibility' (Santer *et al.*, 1996a) – climate data at a single CGCM grid-point represents average conditions over a box with sides of several hundred kilometres in length. But a temperature reconstruction from a single tree-ring site or ice core may record mainly local conditions. This is partially overcome by calibrating the palaeoclimatic series against regional rather than local temperature series (Jones and Briffa, 1996). Of equal concern, however, is the fact that the model noise is an estimate of internally generated variability only, and not of 'total' natural variability (as in palaeo data). A null or inconclusive result in a comparison of model-generated and palaeo-derived noise estimates is thus difficult to interpret.

In summary, there are justifiable concerns over the reliability of both the model-based noise estimates and those obtained from palaeo data. But these concerns do not relieve us of the responsibility of attempting to validate CGCM estimates of multidecadal noise, if we wish to engage in anthropogenic signal detection. As we will show in the following section, model-based noise estimates from the two largest CGCM control runs presently available show quite different partitioning of variance on a wide range of space scales and timescales. Subsequent sections use palaeo data to see which, if either, of these numerical simulations of natural variability has any potential validity.

Model natural variability estimates

This section compares the variability of global mean near-surface air temperature in two control integrations: a 1000-year CGCM experiment recently performed at the Geophysical Fluid Dynamics Laboratory (GFDL; Stouffer *et al.*, 1994) and the first 600 years of what has eventually extended to a 1000-year integration carried out at the Max-Planck Institute for Meteorology (MPI; Cubasch *et al.*, 1994; von Storch, 1994) in Hamburg. A preliminary comparison of the variability in these two integrations was made by Santer *et al.* (1995a). Both models were run with comparable horizontal and vertical resolution. Different flux correction schemes were applied to prevent excessive climate drift (Sausen *et al.*, 1988), but both models still show nonstationary behaviour of global-mean, annually averaged temperature (see Santer *et al.*, 1996b: Figure B1). The residual drift is relatively small in the GFDL run (*c.* 0.1 dg. C/century), but somewhat larger in the MPI control (*c.* 0.2 dg. C/century), primarily due to large changes in sea-ice in the first two centuries of the integration (Santer *et al.*, 1996a). Beyond this time, the run has stability comparable to the GFDL run. In later sections, we investigate sensitivity of the model estimates of natural variability.

The spectra of the detrended, global-mean, annually averaged surface temperature from the observations and the two CGCM integrations are shown in Figure 1. All data sets were detrended prior to spectral analysis. The spectra were computed using the autocorrelation function and smoothed with a 'hanning' filter. The spectra was stable for difference choices of lag and smoothing windows, band widths, etc. The GFDL model spectrum agrees well with the observed spectrum. Unfortunately, the observed record was only 140 years in length, so the observed spectral estimates associated with timescales exceeding 10–20 years are highly uncertain. GFDL has far greater variability than MPI (ECHAM) on timescales less than roughly 65 years, while the reverse applies on longer timescales. It is not fully understood why the two models disagree. The discrepancy between the GFDL and MPI low-frequency power is not attributable to the large initial climate drift in the first few hundred years of the MPI integration since the spectrum is computed with data beginning in year 200, i.e. after the initial drift has ceased and the model has apparently reached an equilibrium climate.

In addition to model differences in variability in global-mean terms, there are also differences in the dominant patterns of variability and in the partitioning of total space-time variance. Patterns of variability were estimated via Empirical Orthogonal Functions (EOFs) of the gridded GFDL and MPI near-surface, annual average air-temperature data. The data were area-weighted prior to calculation of the covariance matrix from which the EOFs were computed. There are striking differences in the spatial patterns of the dominant natural variability modes (Figure 2). In the MPI control run, the variance maxima in the first two spatial EOFs are

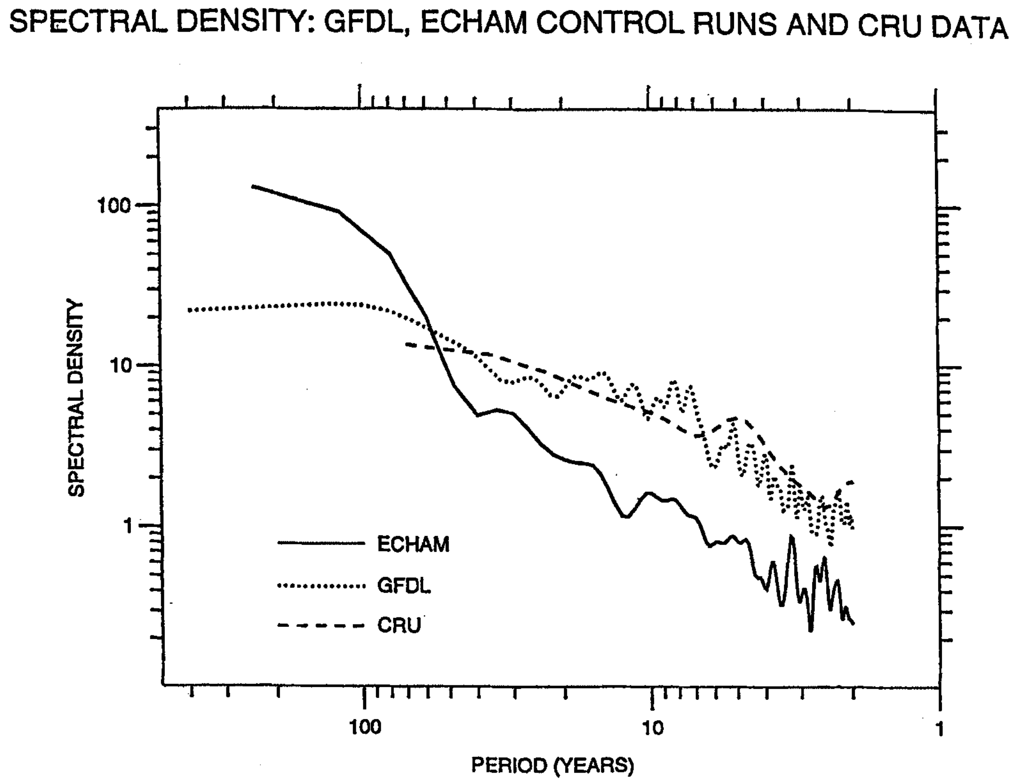

Figure 1 Power spectrum of global mean, annually averaged near-surface air temperature in the long control integrations from coupled general circulation models of the Geophysical Fluid Dynamics Laboratory (GFDL) and Max Planck Institute (MPI), Hamburg. Also shown is a similar calculation based on a 140-year (1854–1993) global near-surface annual average air-temperature record from observations (Climate Research Unit, Jones and Briffa, 1992). All series were detrended prior to spectral analysis.

closely tied to locations where there are long-term changes in sea-ice. In contrast, the dominant modes of variability in the GFDL control run are characterized by largest variability over high-latitude land areas of the northern hemisphere. The eigenvalue spectra of the two experiments (not shown) are also very different, with MPI partitioning twice as much of the total space-time variance in the first two modes than the GFDL model.

These large model-model differences in estimates of the spectrum of natural variability, both in terms of variance levels and large-scale spatial patterns, imply considerable uncertainties in our ability to specify the spectrum of natural variability and subsequently to detect any greenhouse warming signal – even if the space time-evolution of such a signal were perfectly known. Clearly, both models cannot be correct in their estimates of the spectrum of natural variability.

Natural variability in palaeo data

The palaeo data used in this study are those described by Bradley and Jones (1993) and consist mainly of tree-ring reconstructions, historical documentary records (from China and central Europe), ice-melt records, and ice cores. The Bradley and Jones set was augmented with tree-ring data from Tasmania (Cook *et al.*, 1992) and coral data from the Great Barrier Reef (Lough *et al.*, 1996) and the Galapagos (Dunbar *et al.*, 1994). These palaeo proxies are described in Table 1. In the following discussion the data series have been trimmed to cover the common period 1600–1950, or ignored if not of sufficient length. A general description of the eigenstructure of a detailed version of this data set has been made by Mann *et al.* (1994) and the reader is referred to that work for a description of the space-time variability of the palaeo data, results we confirmed by re-analysis.

Several items regarding the data shown in Table 1 and their utility in this study need to be summarized before proceeding with the palaeo-model data intercomparison.

(1) The palaeo records associated with tree-rings had been converted to proxy temperatures by their originators prior to our use of them. These temperatures are most generally representative of the summer season (as assumed by Bradley and Jones, 1993). However the 'W. US density' data are best related to spring/summer season, while the 'Northern Treeline' data are associated with an annual average temperature.
(2) The degree to which the palaeo proxies represent instrumental temperatures was estimated as follows. The observed data from the grid points (or regions) closest to the palaeo proxy sites were isolated for the appropriate season. These observed data were converted to decadal anomaly values and cross correlated with the decadal palaeo data. The results, given in Table 1, show that in almost cases the palaeo proxies represent reasonably well the fluctuations in observed temperature over the last 100 or so years. Detrending both series prior to cross-correlation reduced the size of the correlations by generally small amounts, e.g. 0.10 or less. In some cases, the detrending improved the correlation by a comparable amount. The several series that are not representative of nearby observed temperatures were excluded from the analysis described below, e.g. the Devon Island Ice Cap series and the east China documentary series.
(3) Some of the ice-melt, coral and historical documentary data available to us apparently had not been formally calibrated against an instrumental record. This calibration was carried out here using both the observed, globally gridded air-temperature data set described by Eischeid *et al.* (1996) and that of Jones and Briffa (1992) or the century-long sea-surface

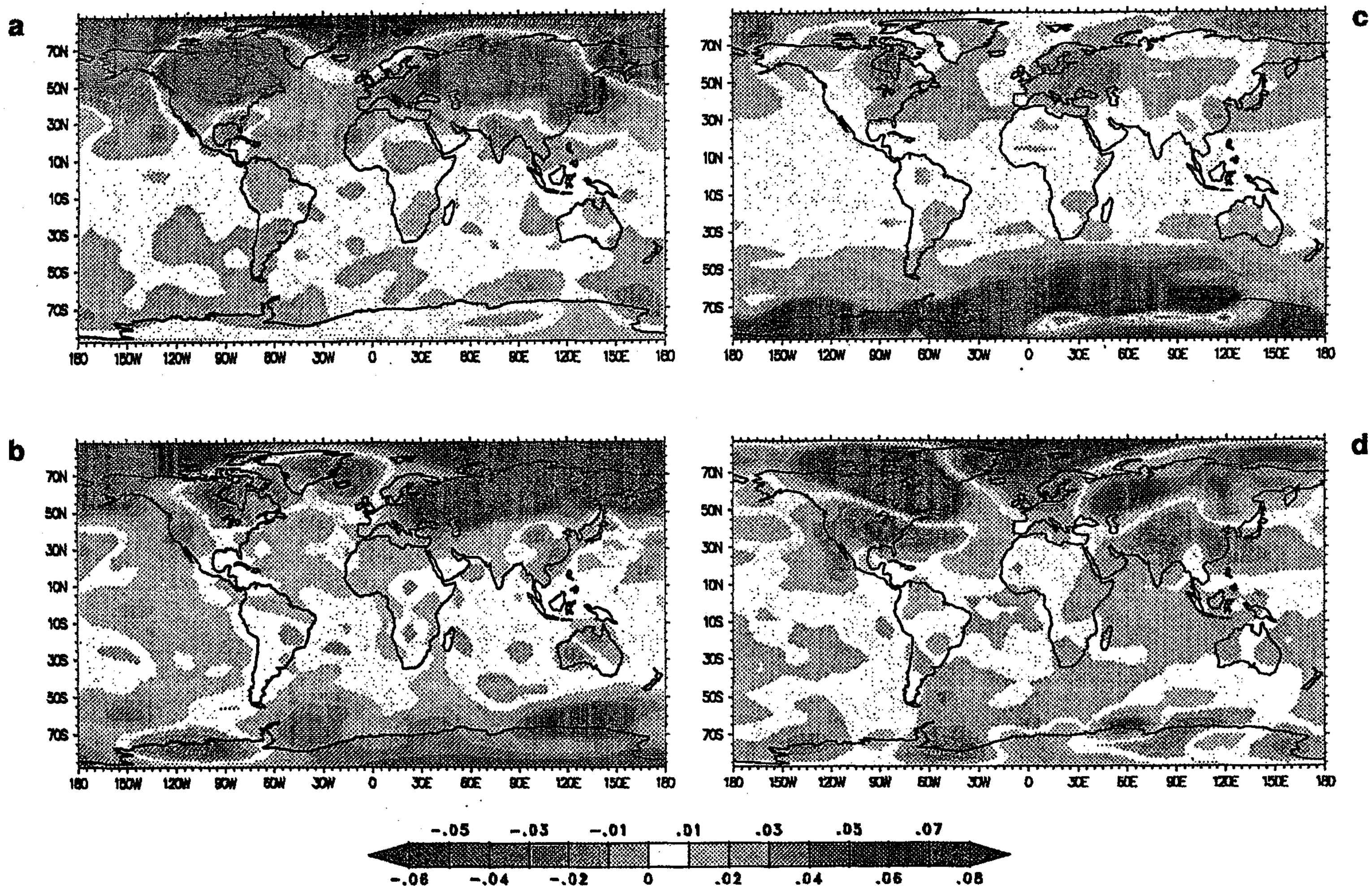

Figure 2 First two empirical orthogonal functions (EOFs) for the GFDL (left panels) and MPI (right panels) control run near-surface air temperature. The EOFs of the GFDL data are based on the first 400 years of the run and account for 6.3% (a) and 4.9% (b) of the variance, respectively. The EOFs of the MPI data are based on years 200–600 of the run and account for 17.7% (c) and 5.7% (d) of the variance, respectively.

Table 1 Palaeo data characteristics

Palaeo site	Data type (units)	Location	Data period	Sampl. interval (Yr)	r	Source (29)
1. Gt Barrier Reef	Coral (iso)	16.7°S, 146.6°E	1615–1982	1	0.66	Lough *et al.*, 1995
2. Galapagos Is.	Coral (iso)	0.2°S, 87.6°W	1607–1981	1	0.70	Dunbar *et al.*, 1994
3. Svalbard	Ice (M)	79.0°N, 15.0°E	1400–1985	1	0.55	Tarussov, 1992
4. N. Scandinavia	Tree (°C)	68.0°N, 22.0°E	1400–1980	1	0.77	Briffa *et al.*, 1992a
5. N. Urals	Tree (°C)	65.0°N, 65.0°E	1400–1969	1	0.80	Graybill and Shiyatov, 1992
6. C. England	Instru. (°C)	51.0°N, 2.0°W	1723–1987	1	0.98	Manley, 1959; 1974
7. C. Europe	Doc (°C)	46.5°N, 8.0°E	1550–1979	1	0.93	Pfister, 1985
8. Agassiz Ice Cap	Ice (M)	80.0°N, 75.0°W	1401–1966	5	0.74	Koerner, 1977
9. Devon Is. Ice Cap*	Ice (M)	70.0°N, 82.0°W	1401–1971	5	0.17	Koerner and Fisher, 1990
10. S. Greenland	Ice (°C)	65.0°N, 45.0°W	1545–1988	1	0.40	Kameda *et al.*, 1992
11. N. Tree-line	Tree (°C)	58.0°N, 95.0°W	1601–1974	1	0.76	D'Arrigo and Jacoby, 1992
12. W. US widths*	Tree (°C)	40.0°N, 112.0°W	1602–1961	1	0.42	Fritts, 1991; Schweingruber *et al.*,
13. W. US density	Tree (°C)	40.0°N, 112.0°W	1600–1982	1	0.85	1991; Briffa *et al.*, 1992b
14. E. China*	Doc	25.0°N, 115.0°E	1400–1980	10	0.13	Wang and Wang, 1990
15. N. China	Doc (°C)	39.0°N, 115.0°E	1400–1980	10	0.86	Wang, 1991a; 1991b
16. S.E. China	Doc (°C)	26.0°N, 118.0°E	1470–1970	10	0.72	Wang *et al.*, 1991
17. Argentina 37–39°S	Tree (°C)	38.0°S, 69.0°W	1500–1974	1	0.59[1]	Villaba, 1990
18. Argentina 41°S	Tree (°C)	41.0°S, 69.0°W	1400–1983	1	0.57[2]	Boninsegna, 1992
19. Lake Johnston	Tree (°C)	42.0°S, 146.5°E	1400–1991	1	0.47[3]	Cook *et al.*, 1991; 1992
20. N. Zealand S. Is.*	Tree (°C)	44.0°S, 171.0°E	1760–1978	1	0.26	Norton and Palmer, 1992

M = ice melt index; doc = documentary; tree = tree-ring; ice = ice core; iso = isotope.
An * following the site name indicates the data were not used in this study. Correlation (r) is between the decadal palaeo data and decadal summer (annual) temperature anomalies from the obesrved, globally gridded near-surface air-temperature data sets (18).
In most cases, the correlations are computed over the last 100 years. [1] correlation over 1921–74, [2] correlation over 1931–83, [3] correlation over 1920–89.

temperature data from the COADS (Slutz *et al.*, 1985) in the following manner: we assumed the three palaeo data types were representative of summer, annual and summer conditions, respectively. Each palaeo time series was normalized to have unit variance over the period 1860–1959. The 100-year record of *observed* data (for the appropriate season) for the data grid point closest to each palaeo site was isolated, and converted to decadal anomalies of summer or annual temperature, as appropriate for the proxy under consideration. The standard deviations of these time series were computed and used to renormalize the associated palaeo time series such that they now had units of degrees Celsius. This procedure assumes the variability during the period of palaeodata is the same as that during the period of observation. The series were then scaled by the correlation (r) between palaeo and observed data, so that the final palaeo series accounted for 100 r^2 % of the variance of the observed data. This, again, assumes a stationary character for both the palaeo and observed data. Bradley and Jones (1993) show an analogous, but slightly different, scaling scheme yields palaeo temperature proxy time series that are an excellent reproduction of observations (see their Figure 7).

(4) Inspection of Table 1 shows that the proxies generally account for only 20–60% of the variance. Hence, the palaeo data apparently underestimates the true levels of natural variability by a factor of 2–5 or more.

Natural variability: models versus palaeo

We have investigated the correspondence between the models and palaeo data in a number of ways. These are summarized briefly below.

(i) Standard deviation ratios

The standard deviations (sigma) of the 350-year, decadal palaeo time series and their counterparts for the first 350 years of the decadal model data for summer (annual average) were computed. The choice of a decadal timescale was driven by the need to maximize the amount of palaeo data available for study; cf. the data sampling intervals shown in Table 1. The ratio of palaeo sigma divided by model sigma was computed next and shown on Figure 3 for the first 350-year segments from the models (essentially the same results were obtained using the second 350-year segment of the model data). In the case of the MPI, the data set was only 600 years and we used the first and last 350 years of this sequence giving 100 years overlap between the two epochs. The ratios are often within a factor of two which suggests roughly the same level of variability between model and data, i.e. the models underestimate the true levels of variability as do the palaeo data. The grand averages of the ratios show the models generally underestimate the palaeo variability by an average factor of 1.5 (GFDL) to 2.1 (MPI) and, hence, underestimate the variance by a factor of about 3–4. The discrepancy is particularly large in the tropical Pacific (Galapagos), since neither model produces a credible ENSO signal, and in southeast China near the Himalayas. The large ratios are due to the models' having very small variability at these locations.

This result is in stark contrast to Figure 1 where the GFDL model seems to reproduce well the spectrum of global mean temperature. It is clear that getting a global integral of temperature correct does not imply that the spatial structure of the signal is also correct. Indeed, the opposite seems to be the case in the current study. The act of integration (to get a global mean) apparently suppresses model limitations. This is yet another demonstration that global mean temperature is not a particularly useful parameter for model validation/climate change detection.

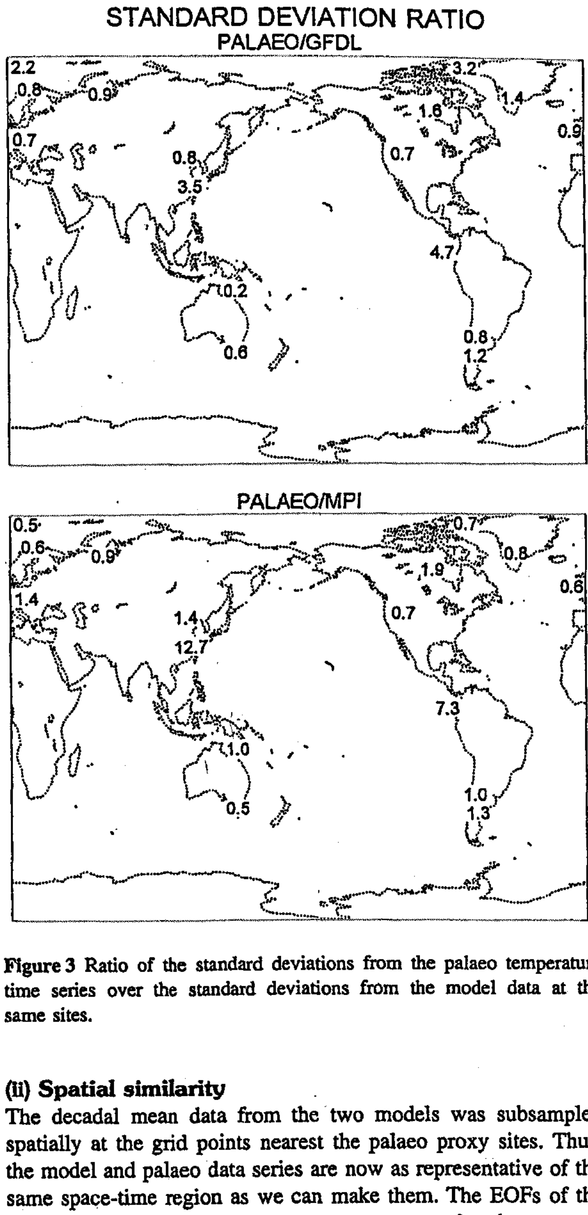

Figure 3 Ratio of the standard deviations from the palaeo temperature time series over the standard deviations from the model data at the same sites.

(ii) Spatial similarity

The decadal mean data from the two models was subsampled spatially at the grid points nearest the palaeo proxy sites. Thus, the model and palaeo data series are now as representative of the same space-time region as we can make them. The EOFs of the covariance matrix of each data set were computed and compared.

The leading EOF of the palaeo data is shown in Figure 4a and accounted for 39% of the palaeo data variance. This eigenmode is statistically distinct from the second and lower modes using the criterion of North *et al.* (1982). It does not replicate well the comparable patterns found for the GFDL and MPI models (Figure 4, b and c; see below). For instance, the MPI model puts the largest variations in the polar latitudes while GFDL favours Europe. However, the explained variance in the first palaeo EOF is about the same as that obtained from the leading EOF of the models. The higher order palaeo EOFs (not shown) agree no better with those obtained from the models. According to (4) above, these results suggest the models are accounting for only 20–60% of the actual natural variability. Since our goal here is to compare the palaeo-model data sets, we refer the reader to the work by Mann *et al.* (1994) for a detailed description of space-time variability of the palaeo data themselves.

The dot products of either of the palaeo-grid model EOFs with the palaeo EOFs is a measure of the similarity of the spatial structure of the data fields. A value of 1.0 would mean an identical spatial structure. The value of the dot products was low, generally less than 0.50 for all combinations of the first ten modes. This means that the spatial structure of the decade to century variability

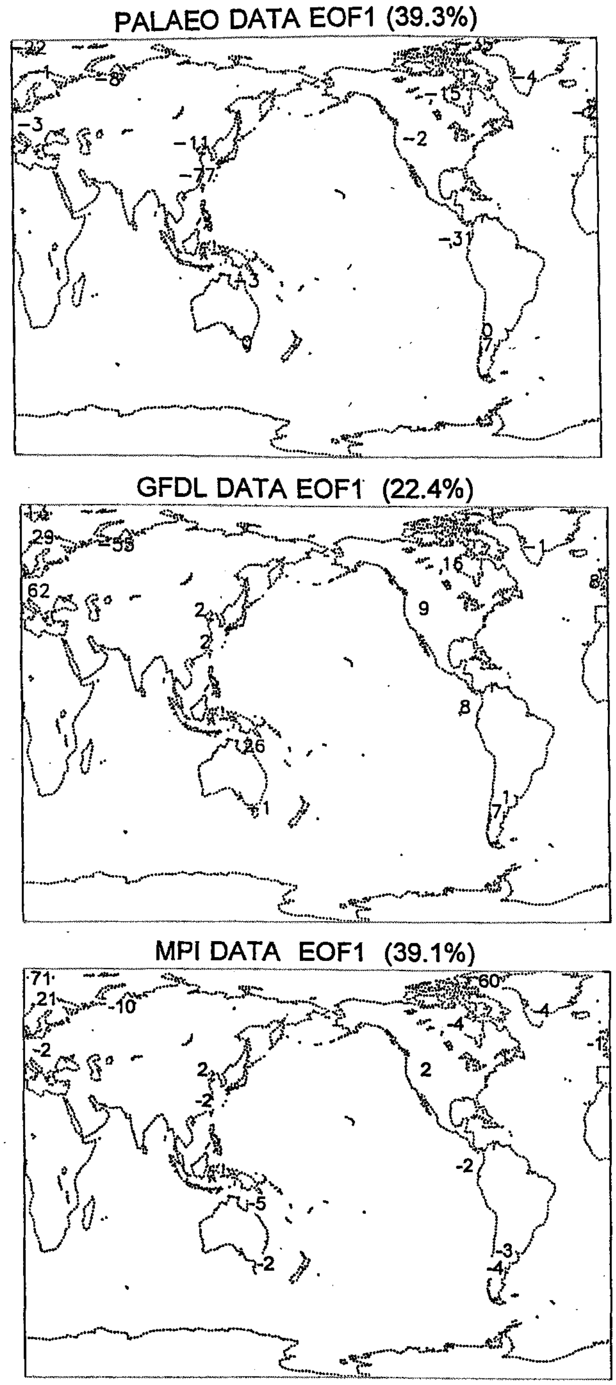

Figure 4 (a) Leading EOF of the 350-year, decadal palaeo data (39.3% of the variance). (b) Ditto for the GFDL data (22.4% of the variance). (c) Ditto for the MPI data (39.1% of the variance). The eigen components have been multiplied by 100 and are positioned at the palaeo site locations.

represented in the palaeo data did not closely resemble that produced by either of the models (but see below). Thus detection methods that rely on pattern recognition techniques (all the modern approaches) have no reliable estimates of the patterns associated with natural variability against which to test for an anthropogenic signal.

(iii) Common basis set

The model summer, or annual, temperature data sets, as appropriate to the particular proxy, were projected onto the palaeo EOFs so that all the data could be compared in a common coordinate system (cf. Barnett and Jones, 1992), thereby overcoming one of the objections to the methodology of (ii) above. Again, decadal data were used to maximize the number of proxies available for analysis and the model data fields were subsampled at the grid points corresponding to the proxy locations. The palaeo principal component for the first EOF is shown in Figure 5. Also shown are the pseudo-principal components obtained by projecting the first 350 years model data onto the leading palaeo EOF. The lower level of variability in the models is obvious and was replicated using the second, independent section of model data. Due to the lack of correlation between model and palaeo EOFs, this situation was found also for all higher modes.

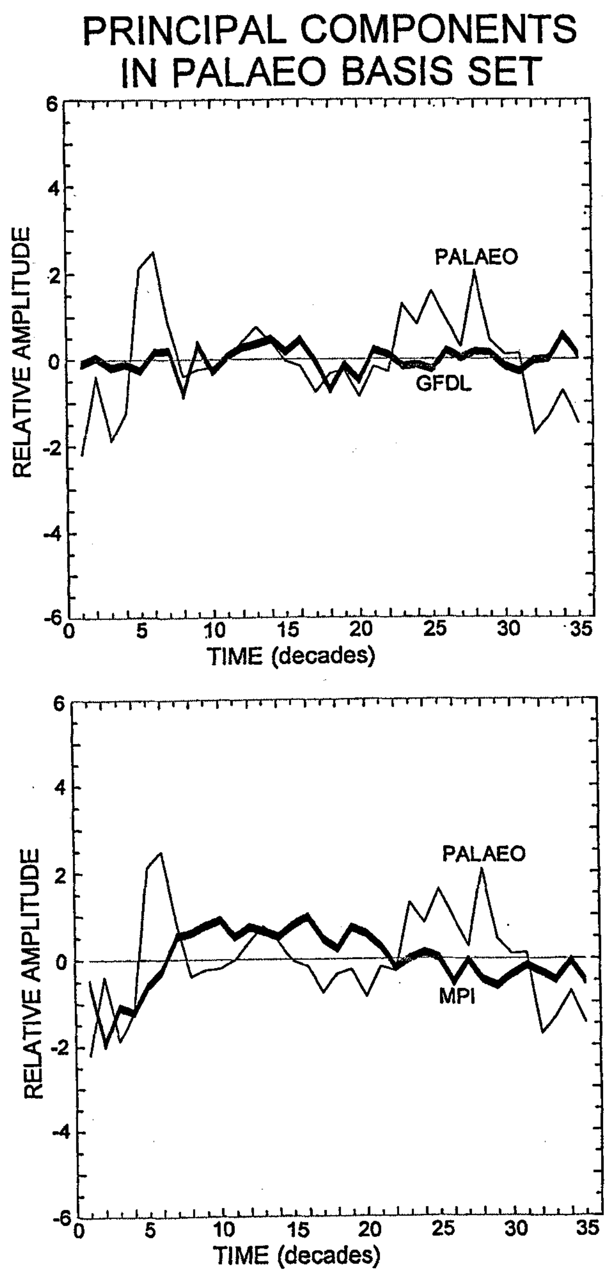

Figure 5 The first principal component (PC) of the palaeo data (lighter line) and the pseudo-principal components (heavy line) of the GFDL (upper) and MPI (lower) obtained by dotting the first 350 years of model data with the leading palaeo EOF. The low amplitude of the pseudo-principal components relative to the palaeo PC shows the models project poorly onto the palaeo EOF. This means the spatial patterns of variability in the models do not resemble well those found in the palaeo data.

The power spectra of these principal components are shown in Figure 6. The spectral analyses were performed on the detrended time series using a Fast Fourier Transform technique with frequency space smoothing via a 'hanning' window applied twice in succession. The general properties of the spectra were essentially unchanged for different reasonable choices of smoothing and effective bandwidths. The spectrum of the palaeo data has a shape that compares favourably with that found by Kutzbach and Bryson (1974). The model spectra demonstrate the reduced variability in the projected models' data sets exists at all frequencies. But the difference becomes larger with decreasing frequency, suggesting a low frequency damping mechanism in the models (e.g. the strong flux correction terms used in both runs, Pierce *et al.*, 1996) or exclusion of some low-frequency physics (e.g. inadequate representation of the global thermohaline circulation). Note, also, in Figure 6 the spectra of the pseudoprincipal components computed from the last 350 years of the model runs carry the same message found from the first 350 years of the runs. The larger low-frequency behaviour of the MPI spectrum for the first 350 years (MPI1) is due to the transient, nonlinear drift mentioned above.

(iv) Sensitivity tests

We tested the robustness of the results obtained in the 'spatial similarity' study – (ii) above – in two ways, as follows.

(a) The palaeo data from the Galapagos and SE China were omitted and the EOF analysis redone on both models and the palaeo proxies. The conclusions regarding the dissimilarity between spatial structures in the palaeo data and the GFDL model did not change. However, the MPI model now demonstrated good EOF pattern correlations with the palaeo EOFS amongst the first three modes ($r = 0.67$ to 0.76). Inspection of the PCs showed that this agreement was due entirely to the initial (nonlinear) drift of the GCM. The apparent good agreement in model-palaeo spatial patterns disappeared when the analysis was repeated on the last 350 years of the MPI control run, a period for which the drift was represented by a simple linear trend.

(b) The analysis described in (iii) above was repeated using palaeo data for the period 1600–1850 to eliminate possible con-

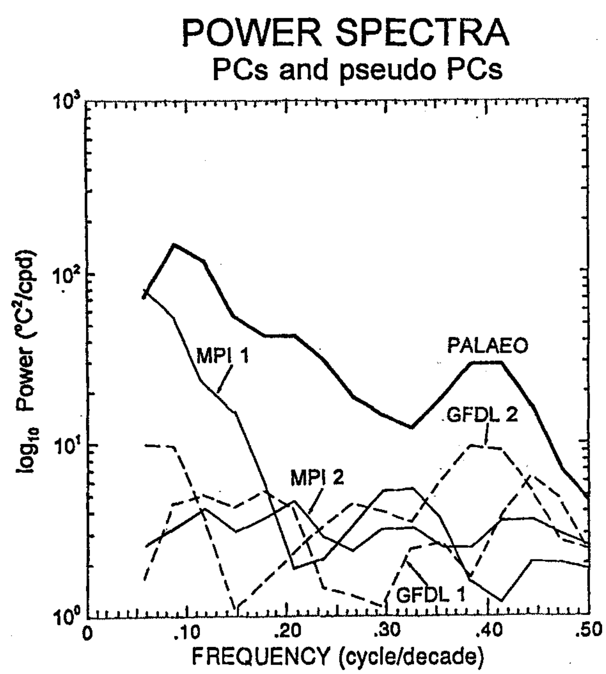

Figure 6 Power spectra of the PCs shown in Figure 5. Also shown are spectra for pseudo-PCs obtained from the last 350 years of each model run denoted by MP12 and GFDL2.

tamination of the results by potential anthropogenic sources. The conclusions were identical to those stated above: the spatial patterns of variability in the palaeo data are substantially different from those produced by the models.

Conclusions

Most of the measures used to compare the model and palaeo data suggest the models are underestimating the levels of decadal-scale natural variability in summer (annual) palaeo temperature proxies. This result is largely driven by a few geographic locations where the models have a particularly low level of decadal variability. Their locations suggest model problems in representing ENSO events, climate conditions around mountain ranges and around large ice concentrations. The discrepancy between the models and the palaeo data is smallest at timescales of two decades (the reciprocal Nyquist frequency of this analysis), but increases substantially with decreasing frequency (increasing timescale). *More importantly for the detection problem is the fact that the spatial patterns of natural variability found in regions from which the palaeo data come do not resemble those produced by either model in the same regions.*

Perhaps it is not surprising that the models underestimate the natural variability, for they do not include effects of volcanoes, possible changes in the solar constant, etc. They also have in common a strong flux correction which may have a stabilizing effect on the model performance (Pierce *et al.*, 1996; Neelin and Dijkstra, 1996). But the palaeo data *also* substantially underestimate the observed variability of the last 100 years. So the fact that the level of variability is within a factor of two between the two types of data is certainly no validation of the model's estimates of natural variability; in fact, just the opposite seems true.

The key message of this paper is that, *if* the palaeo data are reasonably correct and representative of large regions of the planet, then the *current* model estimates of natural variability cannot be used in rigorous tests aimed at detecting anthropogenic signals in the real world. The lack of similarity between the spatial patterns of variability means that simple scaling of the model data will not cure this problem. If the model estimates are used, they are likely to inflate the statistical significance of typical detection metrics by under-representing the air-temperature variance that one should expect in nature. The models also apparently can tell us little about the spatial patterns of climate variability that occur naturally. This result is particularly worrying since the most modern sophisticated detection methods (Hasselmann, 1993; Santer *et al.*, 1996b) try to find predicted spatial patterns of change. All these facts make it difficult to say if *observed* spatial changes in climate are 'normal' or due to anthropogenic effects. One, or both, of these model flaws might bias the results of an objective detection study and lead us to believe confidently that an anthropogenic signal has been found when, in fact, that may not be the case. Expanded, better analyzed palaeo data sets, while containing numerous problems themselves, offer a potential way around these problems.

In closing, it is important to list two major shortcomings of our work. The long CGCM control runs were made with models that were state-of-the-art about five years ago, e.g. the MPI run was done with a first-generation version of the ECHAM model. They are currently running the fourth generation of ECHAM and it is much improved over the earlier version. Another weakness of our study is the small amount of palaeo data available for analysis and the degree to which it is representative of large-scale spatial variability in the surface temperature field. But this situation is changing rapidly with the advent of both national and international efforts to build collections of well-documented palaeo data.

This study clearly needs to be repeated with the newest models and a greatly expanded palaeo data set. Until that time, however, our results should serve as a warning to those anxious rigorously to pursue the detection of anthropogenic effects in observed climate data: the spectrum of natural variability against which detection claims, positive or negative, are made is not well known and apparently not well represented in early CGCM control runs.

Acknowledgements

This work was supported by the US Department of Energy (DOE), Atmospheric and Climatic Research Division, under grant no. DE-FG02-86ER60397 (PDJ), grant no. DE-GF02-89ER69017 (RSB), the National Science Foundation grant no. ATM93-14485 (TPB), DOE Environmental Sciences Division grant no. W-7405-ENG-48 (BDS), by the European Community under grant no. EV5V-CT94-0500 (KRB) and NOAA's Office of Global Programs' Climate Change, Data and Detection program element (TPB).

References

Barnett, T.P. and **Jones, P.D.** 1992: Intercomparison of two different Southern Hemisphere sea level pressure data sets. *Journal of Climate* 5, 93–99.

Boninsegna, J.A. 1992: South American dendroclimatological records. In Bradley, R.S. and Jones, P.D., editors, *Climate since* AD *1500*, London: Routledge, 446–462.

Bradley, R. and **Jones, P.D.** 1993: 'Little Ice Age' summer temperature variations: their nature and relevance to recent warming trends. *The Holocene* 3, 367–76.

Briffa, K.R. and **Jones, P.D.** 1993: Global surface air temperature variations over the twentieth century: Part 2, Implications for large-scale high-frequency palaeoclimatic studies. *The Holocene* 3, 82–93.

Briffa, K.R., Jones, P.D., Bartholin, T.S., Eckstein, D., Schweingruber, F.H., Karlen, W., Zetterberg, P. and **Eronen, M.** 1992a: Fennoscandian summers from AD 500: temperature changes on short and long time scales. *Climate Dynamics* 7, 111–119.

Briffa, K.R., Jones, P.D. and **Schweingruber, F.H.** 1992b: Tree-ring density reconstructions of summer temperature patterns across western North America since 1600. *Journal of Climate* 5, 735–54.

Cook, E., Bird, T., Peterson, M., Barbetti, M., Buckley, B., D'Arrigo, R, and **Francey, R.** 1992: Climate change over the last millenium in Tasmania reconstructed from tree rings. *The Holocene* 2, 205–17.

Cook, E., Bird, T., Peterson, M., Barbetti, M. Buckley, B., D'Arrigo, R., Francey, R. and **Tans, P.** 1991: Climatic change in Tasmania inferred from a 1089-year tree ring chronology of Huon Pine. *Science* 253, 1266–68.

Cubasch, U., Santer, B.D., Hellbach, A., Hegerl, G.C., Hock, H., Maier-Reimer, H., Mikolajewicz, U., Stossel, A. and **Voss, R.** 1994: Monte Carlo climate forecasts with a global coupled ocean-atmosphere model. *Climate Dynamics* 10, 1–19.

D'Arrigo, R.D. and **Jacoby, G.C. Jr.** 1992: Dendroclimatic evidence from northern North America. In Bradley, R.S. and Jones, P.D., editors, *Climate since* AD *1500*, London: Routledge, 296–311.

Delworth, T., Manabe, S. and **Stouffer, R.J.** 1993: Interdecadal variations of the thermohaline circulation in a coupled ocean atmosphere model. *Journal of Climate* 6, 1993–2011.

Dunbar, R.B., Wellington, G.M., Colgan, M. and **Ward Glynn, P.W.** 1994: Eastern Pacific sea surface temperatures since AD 1600. The δ^{18} O record climate variability in the Galapagos corals. *Palaeooceanography* 9, 291–315.

Elscheid, Baker, Karl, T. and **Diaz, H.** 1996: The quality control of long term climatological data using objective analysis. *Journal of Applied Meteorology*, in press.

Fritts, H.C. 1991: *Reconstructing large-scale climatic patterns from tree-ring data*. Tucson: University of Arizona Press.

Graybill, D.A. and **Shiyatov, S.G.** 1992: Dendroclimatic evidence from the northern Soviet Union. In R.S. Bradley and P.D. Jones, editors, *Climate since* AD *1500*, London: Routledge, 393–414.

Hasselmann, K. 1976: Stochastic climate models. Part I, Theory. *Tellus* 28, 473–485.

—— 1993: Optimal fingerprints for the detection of time-dependent climate change. *Journal of Climate* 6, 1957–71.

Hegerl, G.C., von Storch, H., Hasselmann, K., Santer, B.D., Cubasch, U. and **Jones, P.D.** 1996: Detecting anthropogenic climate change with an optimal fingerprint method. *Journal of Climate*, in press.

Jones, P. and **Briffa, K.** 1992: Global surface air temperature variations over the twentieth century. Part I: Spatial, temporal and seasonal details. *The Holocene* 2, 165–79.

—— 1996: What can instrumental records tell us about longer time scale palaeoclimatic reconstructions. In Jones, P., Bradley, R. and Jonzel, J., editors, *Climatic variations and forcing mechanisms of the last 2000 years*. NATC meeting summary, Springer-Verlag, in press.

Kameda, T., Narita, H., Shoji, H., Nishio, F. and **Watanabe, O.** 1992: 450-year summer temperature record from melt feature profile in South Greenland ice core. In: Mikami, T., editor, *Proceedings of the International Symposium on the Little Ice Age Climate*, Tokyo: Dept. of Geography, Tokyo Metropolitan University, 101–107.

Koerner, R.M. 1977: Devon Island icecap: core stratigraphy and palaeoclimate. *Science* 196, 15–18.

Koerner, R.M. and **Fisher, D.A.** 1990: A record of Holocene summer climate from a Canadian high-Arctic ice core. *Nature* 343, 630–31.

Kutzbach, J. and **Bryson, R.** 1974: Variance spectrum of Holocene climatic fluctuations in the North Atlantic sector. *Journal of the Atmospheric Sciences* 8, 1958–1963.

Latif, M. and **Barnett, T.P.** 1994: Causes of decadal climate variability over the North Pacific/North American sector. *Science* 266, 634–37.

—— 1996: Decadal climate variability over the North Pacific and North America: Dynamics and predictability. *Journal of Climate*, in press.

Lough, J.M., Barnes, D.J. and **Taylor, R.B.** 1996: The potential of Great Barrier Reef massive corals for the study of higher-resolution climate variations. In *Variations and forcing mechanisms of the last 2000 years*, Springer-Verlag, in press.

Manabe, S. and **Stouffer, R.J.** 1996: Low frequency variability of surface air temperature in a 1,000 year integration of a coupled ocean-atmosphere model. *Journal of Climate* 9, 376–93.

Manley, G. 1959: Mean temperature of Central England 1698–1952. *Quarterly Journal of the Royal Meteorological Society* 79, 242–61.

—— 1974: Central England temperature data 1941–1979. *Quarterly Journal of the Royal Meteorological Society* 100, 389–405.

Mann, M.E., Park, J. and **Bradley, R.** 1994: Global modes of decadal-to-century-scale climate variability: application of evolutive SVD spectral analysis to globally distributed high-resolution temperature proxy records. *Eos Supplement* 75, 346.

Mehta, V.M. and **Delworth, T.L.** 1996: Decadal variability of the tropical Atlantic Ocean in shipboard measurements and in a global ocean-atmosphere model. *Journal of Climate*, in press.

Neelin, J.D. and **Dijkstra, H.A.** 1996: Ocean-atmosphere interactions and the tropical climatology. Part I: The dangers of flux-correction. *Journal of Climate*, in press.

North, G.R., Bell, T.L., Cahalan, R.F. and **Moeng, F.J.** 1982: Sampling Errors in the Estimation of Empirical Orthogonal Functions. *Monthly Weather Review* 110, 699–706.

Norton, D.A. and **Palmer, J.G.** 1992: Dendroclimatic evidence of northern and central European summer temperatures. In Bradley, R.S. and Jones, P.D., editors, *Climate since* AD *1500*, London: Routledge, 463–82.

Pierce, D., Barnett, T. and **Mikolajewicz, U.** 1996: On the competing roles of heat and fresh water flux in forcing thermohaline oscillations. *Journal of Physical Oceanography*, in press.

Pfister, C. 1985: *Das Klima der Schweiz von 1525–1860*. Bern: Academica Helvetica.

Santer, B.D., Mikolajewicz, U., Bruggemann, W., Cubasch, U., Hasselmann, K., Hoeck, H., Maier-Reimer, E. and **Wigley, T.M.L.** 1996a: Ocean variability and its influence on the detectability of greenhouse warming signals. *Journal of Geographical Research – Oceans*, in press.

Santer, B.D., Taylor, K.E., Wigley, T.M.L., Penner, J.E., Jones, P.D. and **Cubasch, U.** 1996b: Towards the detection and attribution of an anthropogenic effect on climate. *Climate Dynamics*, in press.

Santer, B.D., Wigley, T.M.L., Barnett, T.P. and **Anyamba, E.** 1996c: In IPCC (Intergovernmental Panel on Climatic Change) Report, *Detection of Climate Change, and Attribution of Causes*, in press.

Sausen, R., Barthel, K. and **Hasselmann, K.** 1988: Coupled ocean-atmosphere models with flux corrections. *Climate Dynamics* 2, 145–63.

Schweingruber, F.H., Briffa, K.R. and **Jones, P.D.** 1991: Yearly maps of summer temperatures in western Europe from AD 1750 to 1975 and western North America from 1600 to 1982: Results of a radio-densitometrical study on tree rings. *Vegetatio* 92, 5–71.

Slutz, R., Lubker, S., Hiscoz, J., Woodruff, S., Jenne, R., Joseph, D., Steurer, P. and **Elms, D.J.** 1985: *Comprehensive ocean-atmosphere data set: Release I*. NOAA Environmental Research Laboratories, Climate Research Program, Boulder, CO (NTIS PB86-105723) 268 pp.

Stouffer, R.J., Manabe, S. and **Yinnikov, K.Y.** 1994: Model assessment of the role of natural variability in recent global warming. *Nature* 367, 634–36.

Tarussov, A. 1992: The Arctic from Svalbard to Severnaya Zemlya: Climatic reconstructions from ice cores. In Bradley, R.S. and Jones, P.D., editors, *Climate since* AD *1500*, London: Routledge, 505–16.

Villaba, R. 1990: Climatic fluctuations in northern Patagonia during the last 1000 years as inferred from tree-ring records. *Quarternary Research* 34, 346–60.

von Storch, J-S. 1994: Interdecadal variability in a global coupled model. *Tellus* 46A, 419–32.

Wang, R., Wang, S. and **Fraedrich, K.** 1991: An approach to reconstruction of temperature on a seasonal basis using historical documents from China. *International Journal of Climate* 11, 381–92.

Wang, S. 1991a: Reconstruction of palaeo-temperature series in China from the 1380s to the 1980s. *Wurzburger Geographische Arbeiten* 80, 1–19.

—— 1991b: Reconstruction of temperatures series of North China from 1380s to 1980s. *Science in China Series B* 34, 751–59.

Wang, S. and **Wang, R.** 1990: Seasonal and annual temperature variations since 1470 AD in East China. *Acta Meteorological Sinica* 4, 428–39.

Wigley, T.M.L. and **Barnett, T.P.** 1990: Detection of the greenhouse effect in the observations. In Intergovernmental Panel on Climate Change (IPCC), *Scientific Assessment of Climate Change*, Report Prepared for IPCC by Working Group I, WMO/UNEP, June.