

PCMDI Report No. 21

TOWARDS THE DETECTION AND ATTRIBUTION OF AN ANTHROPOGENIC EFFECT ON CLIMATE

by

Benjamin D. Santer¹, Karl E. Taylor^{1,2}, Tom M. L. Wigley³,
Joyce E. Penner², Philip D. Jones⁴, and Ulrich Cubasch⁵

¹Program for Climate Model Diagnosis and Intercomparison
Lawrence Livermore National Laboratory, Livermore, CA, USA

²Global Climate Research Division
Lawrence Livermore National Laboratory, Livermore, CA, USA

³UCAR Office for Interdisciplinary Earth Studies, Boulder, CO, USA

⁴Climatic Research Unit, University of East Anglia
Norwich, NR4 7TJ, United Kingdom

⁵Deutsches Klimarechenzentrum, Bundesstrasse 55
D-20146 Hamburg, Germany

January 1995

PROGRAM FOR CLIMATE MODEL DIAGNOSIS AND INTERCOMPARISON
UNIVERSITY OF CALIFORNIA, LAWRENCE LIVERMORE NATIONAL LABORATORY
LIVERMORE, CA 94550

DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This is an informal report intended primarily for internal or limited external distribution. The opinions and conclusions stated are those of the author and may or may not be those of the Laboratory.

This report has been reproduced
directly from the best available copy,

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information
P.O. Box 62, Oak Ridge, TN 37831
Prices available from (615) 576-8401, FTS 626-8401

Available to the public from the
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Rd.,
Springfield, VA 22161

Abstract

It has been hypothesized recently that cooling caused by anthropogenic sulfate aerosols may be obscuring a warming signal associated with changes in greenhouse gas concentrations. Here we use results from model experiments in which sulfate and carbon dioxide have been varied individually and in combination in order to determine whether the simulated surface temperature change patterns are increasingly evident in observed records of temperature change.

We use centered [$R(t)$] and uncentered [$C(t)$] pattern correlation statistics in order to compare observed time-evolving surface temperature change patterns with the model-predicted equilibrium signal patterns. We show that in the case of temperature signals from the "CO₂-only" and "sulfate-only" experiments, the $C(t)$ statistic essentially reduces to a measure of observed global-mean temperature changes, and cannot be used to uniquely attribute observed climate changes to a specific causal mechanism. For the signal from the experiment with combined CO₂ / sulfate aerosol forcing, $C(t)$ provides information on pattern congruence, but trends in $C(t)$ are difficult to interpret without decomposing the statistic into pattern similarity and global-mean change components. We therefore focus on $R(t)$, which is a more useful statistic for discriminating between forcing mechanisms with different pattern signatures but similar rates of global mean change.

Our results indicate that in summertime (JJA) and fall (SON), the pattern of near-surface temperature change in response to combined sulfate aerosol/CO₂ forcing shows increasing similarity with observed changes over the last 50 years. The results from the individual CO₂-only and sulfate-only experiments suggest that at least some of this increasing spatial congruence is attributable to areas where the real world has cooled.

In the sulfate-only experiment, we find that the location of the maximum response differs from the location of the maximum forcing. This illustrates the importance of the atmospheric general circulation in modulating the response to regionally-localized forcing, and points towards possible problems with sulfate aerosol detection studies that do not involve an atmospheric model, and that compare instead observed temperature changes with observed patterns of sulfur emissions or the sulfate aerosol distribution predicted by a sulfur chemistry model.

In the absence of reliable information on the magnitude and spatial characteristics of long time scale natural variability in the real world, we use data from multi-centennial control integrations performed with two different coupled atmosphere-ocean models to estimate the sampling distributions of trends in $R(t)$ and $C(t)$ on time scales of 10- to 50 years. For the combined sulfate aerosol/CO₂ experiment, the 50-year $R(t)$ trends for the JJA and SON signals are significant relative to the trends obtained in the absence of external forcing. Results are robust in that they do not depend on the choice of control run used to estimate natural variability noise properties. The $R(t)$ trends for the CO₂-only signal are not significant in any season.

The caveats regarding the signals and natural variability noise which form the basis of this study are numerous. Nevertheless, we have provided first evidence that both the largest-scale (global-mean) and smaller-scale (spatial anomalies about the global mean) components of a combined CO₂ / anthropogenic sulfate aerosol signal are identifiable in the observed near-surface air temperature data. If the coupled-model noise estimates

used here are realistic, we can be highly confident that the anthropogenic signal which we have identified is distinctly different from natural variability noise. The fact that we have been able to detect the detailed spatial signature in response to combined CO₂ and sulfate aerosol forcing, but not in response to CO₂ forcing alone, suggests that some of the regional-scale background noise (against which we were trying to detect a CO₂-only signal) is in fact part of the signal of a sulfate aerosol effect on climate. The large effect of sulfate aerosols found in this study demonstrates the importance of their inclusion in experiments designed to simulate past and future climate change.

1 Introduction

Most previous greenhouse-gas (GHG) detection studies employing the so-called "fingerprint" strategy introduced by Madden and Ramanathan (1980; see also MacCracken and Moses, 1982) have used some form of pattern correlation statistic to compare the pattern of a model-predicted GHG signal with the time history of observed patterns of near-surface temperature changes (e.g. Barnett, 1986; Barnett and Schlesinger, 1987; Santer et al., 1993). The signal pattern in such studies is usually taken from an equilibrium CO₂-doubling experiment, or towards the end of an experiment with a time-dependent GHG increase at a point where the response pattern is relatively stationary.

The strategy in this method is to search for a long-term, positive trend in the pattern correlation statistic, which would indicate an increasing expression of the GHG signal in the observations. Previous fingerprint studies have employed either uncentered statistics, in which the searched-for signal consists of both a pattern and the global-mean change (e.g. Barnett and Schlesinger, 1987; Hegerl et al., 1994), and/or uncentered statistics, in which the global-mean change is removed and the signal is simply a spatial anomaly pattern (Santer et al., 1993). A recent investigation making use of an uncentered statistic reached the conclusion that a model-predicted GHG signal was identifiable with a high level of confidence in the observed data (Hegerl et al., 1994). This and previous studies have shown that the global-mean change is an important component of a GHG signal. A large body of earlier work has also shown that observed changes in global-mean annually-averaged near-surface air temperature are significant relative to various statistical- and model-based estimates of natural variability noise (e.g., Wigley et al., 1989; Wigley and Raper, 1990, 1991a,b; Bloomfield and Nychka, 1992; Karl et al., 1991; Stouffer et al., 1994).

Such studies have not directly addressed the issue of establishing an unambiguous link between changes in some external forcing factor and changes in observed climate. This is the attribution issue. It is possible that different external forcing mechanisms (or combinations of external forcings and natural variability) can give rise to similar changes in the global-mean state. Since changes in the value of an uncentered statistic largely reflect a change in the global mean, such analyses provide little help in discriminating between different mechanisms that could have produced such a change.¹ The studies cited above suggest that

¹ Unless the change in the global mean is so large that it cannot be explained by non-anthropogenic forcing

we might attach high confidence to the statement that the observed changes in global-mean temperature over the past century are significant relative to current 'best estimates' of the magnitude of natural variability noise. However, the same investigations do not allow one to attach high confidence to the statement that the observed changes are solely attributable to an enhanced greenhouse effect.

Our level of confidence in attributing observed changes to anthropogenic influences would be increased if we could demonstrate that even small-scale spatial features of a model-predicted anthropogenic signal showed a correspondence with observed changes. This is where centered statistics are useful, since they focus on anomalies about the global mean. The only previous study which employed a centered statistic to search observed records of near-surface air temperature for model-predicted GHG signals failed to show any meaningful multi-decadal positive trends in the measure of pattern correspondence (Santer et al., 1993). This negative result has a number of possible explanations, such as errors in the predicted GHG signal pattern, or masking of regional-scale features of the signal by low-frequency natural variability and/or other forcing factors (anthropogenic sulfate aerosols, volcanic aerosols, solar variability, etc.).

Until recently, it has not been possible to investigate the plausibility of these alternative explanations in a quantitative way. Model experiments recently performed by Taylor and Penner (1994; henceforth TP), however, may help to clarify whether the detection and attribution of an anthropogenic effect on climate can be facilitated by incorporating the climatic effects of anthropogenic sulfate aerosols. Sulfate aerosols arise from the SO₂ emitted by fossil fuel combustion, industrial activities, and biomass burning. Such aerosols are likely to have caused some degree of regional-scale cooling (Wigley, 1989), both directly through clear-sky radiative forcing (reflection of incident solar radiation) and indirectly due to changes in cloud brightness (Wigley, 1989, 1991; Charlson et al., 1991, 1992; Kiehl and Briegleb, 1993; Taylor and Penner, 1994; Charlson and Wigley, 1994).

The TP integrations used an atmospheric general circulation model (AGCM) coupled to a tropospheric chemistry model to investigate the climate response to forcing by anthropogenic sulfate aerosols and CO₂. Four integrations were performed: a control run with no anthropogenic sulfate aerosols and nominal pre-industrial CO₂ concentrations, a 'sulfate-only' experiment with present-day sulfur emissions, a 'CO₂-only' experiment with mechanisms and/or natural variability.

near-current atmospheric CO₂ concentrations, and a 'combined forcing' experiment with present-day sulfur emissions and CO₂ levels. The experiments indicated that the sulfate-induced cooling pattern was very different from the (negative of the) pattern of greenhouse warming. More importantly, the response pattern in the combined forcing experiment was generally dissimilar to the response patterns in the sulfate-only and CO₂-only experiments. These results strongly indicate that the combination of sulfate and CO₂ forcing yields a signal pattern substantially different from the greenhouse warming pattern typically produced by AGCMs coupled to mixed-layer oceans (e.g. Manabe and Stouffer, 1980; Hansen et al., 1984; Schlesinger and Mitchell, 1987; Washington and Meehl, 1989) or fully-coupled A/OGCMs (Stouffer et al., 1989; Cubasch et al., 1992; Meehl et al., 1993). If CO₂ and anthropogenic sulfate aerosols have both affected climate, it would seem more appropriate to compare observed data with the signal pattern from the combined sulfate/CO₂ experiment than with the CO₂-only response pattern.

The TP integrations represent a first attempt to simulate the three-dimensional climate response to anthropogenic sulfate aerosol forcing with an AGCM coupled to a model of tropospheric sulfate chemistry. The only comparable studies at present are those by Roeckner et al. (1994) and Mitchell et al. (1994). These investigations also considered the temperature response to a combination of GHG and anthropogenic sulfate aerosol forcing, but the annual mean sulfate aerosol burden was prescribed and the direct radiative effect of the aerosol was parameterized by changing the surface albedo. Both studies did not allow changes in climate to influence the aerosol distribution, whereas in the TP study two-way interaction is possible between climate and aerosols.

To date, no attempt has been made to search observed records of near-surface temperature changes for a GCM-predicted, two-dimensional temperature-change pattern due to sulfate aerosols or some combination of sulfate aerosol and CO₂ forcing. Investigations of the possible impact of sulfate aerosols on observed temperature data have thus far been restricted to analysis of observed changes in areas where sulfate aerosol emissions and/or forcing is likely to have been large (Wigley et al., 1992; Engardt and Rodhe, 1993; Hunter et al., 1993; Karl et al., 1994) or to a visual comparison of observed temperature change patterns and patterns of vertically-integrated aerosol concentration predicted by a chemical-transport model (Engardt and Rodhe, 1993).

In the present study we extend the pattern correlation analysis of Santer et al. (1993)

and determine whether the near-surface temperature signals in the TP sulfate-only, CO₂-only and combined forcing experiments are increasingly evident in the observed data. The temperature signals used in our investigation are taken exclusively from equilibrium response experiments. Our strategy of searching the time-varying instrumental temperature records for an equilibrium signal pattern, with no time information other than a seasonal dependence, requires some explanation.

There are at least three reasons for our focus on time-independent signals. The first reason is data availability: at present, no A/OGCM has been forced by time-dependent changes in both anthropogenic sulfate aerosols and GHG concentrations. When such an integration is performed, it will require a pre-industrial starting time² and a coupled model which has had a spin-up period of sufficient length to attain a quasi-stationary state.³

Second, even if we assumed that a particular A/OGCM contained all the physics necessary to accurately simulate the climate system's response to GHG and/or sulfate aerosol forcing, a single transient experiment performed with such a model and observed SO₂ emissions and GHG concentration changes would be unlikely to provide a reliable estimate of the true, underlying space-time signal in response to the combined anthropogenic forcing. This is due to the fact that the A/OGCM generates its own complex internal variability as a result of interactions between the atmosphere and ocean, unrelated to changes in greenhouse gases, sulfate aerosols, or other external forcings. As shown by Cubasch et al. (1994), an ensemble of transient greenhouse warming integrations with identical time-evolving GHG forcing but small differences in the initial boundary conditions can yield rather different solutions for the time-evolving response patterns (at least on timescales of up to 40-50 years). This result illustrates that different manifestations of the model's natural variability noise are superimposed on the true, underlying GHG signal, so that a reliable estimate of the signal's initial evolution is likely to require averaging over a number of different realizations. Initial condition uncertainty will be even more of an issue in obtaining a reliable estimate of the true space-time signal in response to combined GHG/sulfate aerosol forcing, since the signal-to-noise ratios are probably much lower than for the case of GHG forcing alone.⁴

²This is necessary in order to alleviate the so-called 'cold start' error; see e.g., Fichefet and Tricot, 1992; Hasselmann et al., 1993.

³In order to avoid the convolution of initial equilibration phenomena with the the predicted anthropogenic climate change signal; see Santer et al., 1994a.

⁴This is due to the fact that we are looking for a signal that results from partially-cancelling positive and

Third, there are large uncertainties in the past history of anthropogenic aerosol emissions, and even larger uncertainties in the magnitude and pattern of aerosol forcing changes. These translate into further uncertainty in the space-time evolution of the climate-change signal.

It is for these reasons that we focus on equilibrium signal patterns. This strategy makes the implicit assumption that the true time-varying signal pattern (if such could be obtained) varies little in time and is highly similar to the equilibrium signal pattern. This assumption, in turn, requires that historical differences in the patterns of SO_2 emissions (and in the relative magnitudes of GHG and aerosol forcing) are small. At present, we have no choice other than to search for an equilibrium GHG/sulfate aerosol response pattern in the observed data, and it may in fact be difficult to define a reliable transient GHG/sulfate aerosol signal.

The structure of this paper is as follows. Section 2 provides a brief description of the TP numerical experiments. In Section 3 we discuss the geographical distributions of seasonally- and annually-averaged changes in near-surface temperature in the three forcing experiments and consider the statistical significance of changes in means using univariate (grid-point) t-tests. We also employ spatial correlations to measure the pattern similarity of various quantities in the response experiments. In Section 4, we first define centered and uncentered pattern similarity statistics, $R(t)$ and $C(t)$, and briefly consider their relative merits in the context of detection and attribution of observed climate change. Section 5 then employs these statistics to search the observed record of near-surface temperature changes for the seasonal and annual surface temperature signals from the three TP (1994) perturbation experiments. The question of whether the trends in our pattern similarity measures are significant is addressed in Section 6. In the absence of reliable information on the magnitude and spatial characteristics of decadal- to century-timescale natural variability in the real world, we employ data from two long control integrations performed with fully-coupled A/OGCMs to estimate the sampling distributions of ‘unforced’ trends in $R(t)$ and $C(t)$ on timescales of 10- to 50 years and to assess the significance of observed trends in these statistics. A short summary and conclusions are given in Section 7.

negative forcings.

2 Experiments

The experiments discussed here have been described in detail by TP. The integrations were performed with the GRANTOUR tropospheric chemistry model developed at Lawrence Livermore National Laboratory (Walton et al., 1988), coupled to the Livermore version of the National Center for Atmospheric Research Community Climate Model (NCAR CCM1; Taylor and Ghan, 1992). The CCM1 atmospheric general circulation model in turn was coupled to a 50-meter mixed-layer ocean model with prescribed meridional heat flux. CCM1 has 12 layers in the vertical, and was run with a horizontal resolution of ca. 4.5° latitude \times 7.5° longitude. The GRANTOUR tropospheric chemistry model is a Lagrangian trace species model, which simulates the transport, transformation and removal of various sulfur species (Penner et al., 1994a). Only the direct effects of sulfate aerosols are considered. For further details of the GRANTOUR/CCM1 experimental configuration, refer to TP.

Four integrations were performed: a control run (CTL) with nominal pre-industrial CO_2 levels (270 ppmv; the standard IPCC value is 278 ppmv, Schimel et al., 1994) and no anthropogenic sulfur emissions, a sulfate-only experiment (S) with near-present-day anthropogenic sulfur emissions (prescribed according to Spiro et al., 1992, and Benkowitz, 1982) and near-pre-industrial CO_2 concentrations, a CO_2 -only experiment (C) with no sulfur emissions and nominal present-day CO_2 levels (345 ppmv; c.f. the 1990 value of 354 ppmv given by Schimel et al., 1994), and a combined forcing experiment with near-present-day sulfur emissions and CO_2 . For the purposes of this study, each of the original TP integrations was extended by at least 10 years, and temperature-change signals were computed using samples from the last 20 years of each simulation (following at least a 10-year spin-up period).

3 Model Results

We compare the surface temperature responses in the three perturbation experiments in various ways – using maps of the geographical distribution of temperature change, pattern correlations, and univariate t-tests (e.g., Wigley and Santer, 1990; Santer and Wigley, 1990). For comparing patterns, we computed centered (spatial mean removed) spatial correlations, here denoted by R (see Santer et al., 1993). We also computed an area-weighted form of R .

Results for the unweighted and area-weighted forms of R are similar, and in the following the discussion is restricted to the area-weighted form only. Univariate t-tests were used to determine whether (and where) the grid-point means were significantly different in the control and response experiments.

3.1 Sulfate-Only Experiment

Figures 1, 2, and 3 show the geographical distributions of seasonally- and annually-averaged changes in near-surface temperature in the three TP perturbation experiments. In the S experiment, temperature changes in DJF, JJA and the annual average are negative at virtually all grid-points. The maximum cooling occurs over the Norwegian Sea (ca. -7°C) in winter and to the east of the Weddell Sea (-6°C) in summer.

In both DJF and JJA, the location of the maximum response differs from the location of the maximum forcing; the spatial pattern correlations between radiative forcing and response are only 0.02 in DJF and 0.36 in JJA (c.f. Figures 1, 2, and Figure 4; also Table 1 and TP). The higher correlation in JJA reflects the fact that there is some spatial congruence between forcing and response where both are large over Western Europe. In contrast, the DJF response maximum in the Norwegian Sea is not congruent with a maximum in the forcing. A similar displacement between forcing and response maxima is suggested by the results of Roeckner et al. (1994).

Although there are substantial uncertainties associated with the forcing and response patterns simulated in both the Roeckner et al. and TP studies, the results illustrate the importance of the atmospheric general circulation in 'modulating' the response to regionally-localized forcing. Detection studies that use as their signal the sulfate aerosol distribution or aerosol forcing predicted by a sulfur chemistry model (e.g., Wigley et al., 1992; Engardt and Rodhe, 1993), or pattern of observed changes in sulfur emissions (Karl et al., 1994) may therefore yield misleading results.

The temperature changes noted above are large compared with changes observed this century, mainly because they are equilibrium changes. While these magnitudes may be debatable, the actual values are not relevant in the present study, which focusses on the patterns of change - i.e., the spatial distribution of relative changes.

Note that the maximum Northern Hemisphere cooling occurs in DJF, even though the maximum forcing is in JJA and the DJF forcing is relatively small (c.f. Figures 1a, 2a, and Figure 4). In JJA, the maximum response is in the Southern Hemisphere, even though the forcing is largely in the Northern Hemisphere. There are several explanations for these results. The fact that the Southern Hemisphere demonstrates a substantial response to a forcing primarily confined to the Northern Hemisphere (c.f. Figures 2a, 3a and 4) is an indication of the ability of the atmosphere to exchange heat and momentum between hemispheres in an efficient way. The generally stronger response in the winter hemisphere is in part due to the fact that the effect of changes in sea-ice extent on surface temperature is largest at this time of year (when surface inversions are possible).

A univariate t-test⁵ reveals that the changes in mean state are highly significant in the sulfate-only experiment (Table 2 and Figure 5). 86.5% - 98.8% of the total number of grid-point tests performed showed significant differences in temperatures between the experiment and control (at the $\alpha = 0.01$ significance level). It is clear from these results (even without resorting to multivariate significance tests; see Preisendorfer and Barnett, 1983; Wigley and Santer, 1990; Santer and Wigley, 1990) that the overall (global) differences in means between the sulfate-only and control experiments are highly significant.

3.2 CO₂-Only Experiment

The spatial patterns of seasonally- and annually-averaged near-surface temperature changes in the CO₂-only experiment are very similar to those obtained in equilibrium CO₂-doubling integrations with comparable AGCM/mixed-layer ocean experimental configurations (Manabe and Stouffer, 1980; Hansen et al., 1984; Schlesinger and Mitchell, 1987; Washington and Meehl, 1989). The warming is spatially coherent (Figures 1b, 2b, 3b). The DJF and JJA temperature change patterns show the standard picture of equator-to-pole amplification in the winter hemisphere, a feature associated with poleward retreat of the sea-ice margin. This pattern similarity occurs despite the comparatively low level of the forcing - the CO₂-only experiment was performed with an atmospheric CO₂ concentration change of only 75 ppmv (from 270 to 345 ppmv), in contrast to the ca. 300-330 ppmv changes (from ca. 300-330 ppmv to 600-660 ppmv) commonly used in step-function CO₂ doubling experiments.

⁵All t-tests were performed using 20-year samples from the control run and the perturbation experiments.

The overall maximum temperature increases tend to occur in locations where maximum cooling occurred in the sulfate-only experiment, i.e., in the Norwegian Sea in DJF (ca. 7°C) and in the Ross and Weddell Seas in JJA (ca. 7-8°C). The annual average changes do not show the same degree of hemispheric symmetry commonly found in equilibrium doubling experiments with mixed-layer models (Schlesinger and Mitchell, 1987). This is due to the smaller percentage reduction in sea-ice coverage in the Northern Hemisphere – the reduction in ice-coverage in the Southern Hemisphere is nearly three times larger (see Table 1 of TP). The seasonal and annual warming patterns have a land-sea contrast component, with larger changes over land areas, as in recent transient experiments with fully-coupled A/OGCMs (Cubasch et al., 1992; Santer et al., 1994b). Unlike these recent experiments, the land warming maxima are not concentrated in desert areas with low evaporative cooling.

As in the sulfate-only case, the response pattern differs markedly from the pattern of the forcing, and is negatively correlated with the latter ($R = -0.51$ in DJF, $R = -0.52$ in JJA; see Table 1). This inverse relationship is due to the different zonal structures of forcing and response fields – while the response is a maximum at high latitudes in both hemispheres, the forcing peaks at low latitudes (c.f. Figures 1b, 2b and 4c, 4d).

The univariate t-test results indicate that the surface temperature signal in the CO₂-only experiment is highly significant in all seasons and in the annual average (Table 2). Changes in the CO₂-only experiment show consistently higher percentages of grid-points with significant differences in means than in the sulfate-only experiment, primarily due to the larger global mean temperature change in the former experiment (+1.45°C versus -1.19°C, respectively).⁶ Note that significant changes in annually-averaged temperature cover larger percentage areas than seasonal temperature changes, since the averaging reduces noise.

⁶Note that in TP, the climate sensitivity of the model for a doubling of CO₂ was given as 6.4°C, a result which was based on 10-year samples only. The results which we present here are now based on the final 20 years of much longer integrations (30-50 years). The climate cooled by several tenths of a degree in the CO₂-only integration (before stabilizing towards the end of the integration), yielding a much lower estimate of the climate sensitivity: 4.8°C.

3.3 Combined Sulfate/CO₂ Experiment

Unlike the sulfate-only and CO₂-only integrations, the patterns of temperature change in the combined sulfate/CO₂ experiment are characterized by spatially-coherent regions of both warming and cooling (Figures 1c, 2c and 3c). Temperature decreases are restricted largely to the Northern Hemisphere, which is where most of the radiative forcing associated with sulfate aerosols occurs (Figure 4). The largest decreases are over the Norwegian Sea in DJF (ca. -2°C) and over south-eastern Europe in JJA (ca. -2°C). Warming maxima are in the Ross and Weddell Seas in JJA (ca. 4-8°C), and over Greenland, Labrador, the Sea of Okhotsk, and a small area of Antarctica in DJF (ca. 2°C).

The warming and cooling maxima in all seasons are considerably reduced relative to the respective maxima in the sulfate-only and CO₂-only experiments. In DJF, for example, the large temperature changes in the Norwegian Sea in the S and C experiments (Figures 1a, 1b) are considerably reduced in the SC integration (Figure 1c). While the DJF response patterns in the SC and S integrations show some spatial correspondence ($R = 0.46$), the SC and C response patterns are uncorrelated ($R = 0.01$; see Table 3). In JJA, however, the SC response pattern is very similar to the CO₂-only temperature change pattern ($R = 0.79$), and is negatively correlated with the sulfate-only response pattern ($R = -0.26$). This is due to the fact that the offsetting effects of sulfate in the regions of maximum temperature response to CO₂ forcing are less in JJA than in DJF (c.f. Figures 1a,b and 2a,b). The pattern similarity between C and SC and S and SC clearly depends on the relative magnitudes of the CO₂ and aerosol forcing.

The univariate t-tests (Table 2) indicate that the fractions of the globe with significant differences in means (SC versus CTL) are consistently lower than in either the S or C experiments. This is attributable to compensating warming and cooling responses over large areas of the Northern Hemisphere in both seasons and in the annual average, resulting in large areas with relatively small changes in the mean state. Large, spatially-coherent regions with differences in means significant at the 1% and 5% levels are generally restricted to Southern Hemisphere oceans and low-latitude Northern Hemisphere ocean areas (Figure 5). In DJF, the only land areas showing evidence of a significant response are over Greenland, southern Europe, Antarctica, and areas of South America and Africa. In JJA, significant responses are found over Alaska, south-eastern Europe, and areas of Australia and Antarctica.

The limited statistical significance of the response over large areas of the Northern Hemisphere is not an artefact of the length of the control and response experiments. For the ‘present-day’ sulfate and CO₂ forcing levels stipulated in the TP SC experiment, large areas of the Northern Hemisphere with little or no change in the mean state are a fundamental property of the simulated response pattern (at least for the model used here).

It is interesting to note that the response patterns for the SC integration are, in a purely qualitative sense, more similar to observed patterns of seasonal temperature change (see Jones et al., 1991; Folland et al., 1992; Parker et al., 1994) than the response patterns in the individual S and C experiments. We will consider this issue further in Section 5.1, which presents quantitative measures of observed versus simulated pattern similarity.

4 Pattern Similarity Statistics

In this section we introduce the pattern correlation statistics, $R(t)$ and $C(t)$, which we subsequently use for comparing model and observed spatial patterns of temperature change. It is assumed in the following that we are dealing with seasonally- or annually-averaged data. The terminology is similar to that used by Santer et al. (1993).

4.1 Definition and Computation of Pattern Similarity Statistics

We first compute seasonal- and annual-mean temperature changes in the TP control run and response experiments. In each of the three TP response experiments (S, C, and SC), the temperature-change signal is defined as

$$\Delta M(x) = \overline{M_{\text{EXP}}}(x) - \overline{M_{\text{CTL}}}(x) \quad (1)$$

where M denotes model data, with the subscripts EXP and CTL identifying output from one of the TP response experiments and control integration, respectively. The index x is a discrete variable running over space (grid-points), with $x = 1, \dots, n$. The overbars in $\overline{M_{\text{EXP}}}(x)$ and $\overline{M_{\text{CTL}}}(x)$ indicate time averages, here computed using 20-year samples of

experiment and control data. We stress that these signal patterns have no time-dependence (other than a seasonal dependence), since the TP integrations are equilibrium response experiments with no interannual changes in the forcing.

The observed data, $D(x, t)$, consist of monthly-mean, land-based surface air temperatures and sea-surface temperatures from the combined land-ocean data set described by Jones et al. (1991), spanning the interval 1854-1993. The data are in the form of anomalies relative to the mean over 1950-79. Observed data were processed in the following way. We first define a temporally-smoothed reference state $\overline{D}_1(x)$ centered on the year t_0 as

$$\overline{D}_1(x) = \sum_{v=-q}^{+q} D(x, t_0 + v)W(v) \quad (2)$$

where $W(v)$ are the normalized symmetric weights for a p -term Gaussian filter centered at t_0 (with $p = 2q + 1$; we use $p = 13$ below). As an example, if $t_0 = 1954$, the reference climate is the filtered mean over the years 1948-1960 (we will also consider other values of t_0 in Appendix A in order to examine the sensitivity of results to choice of reference period). We then define anomalies relative to $\overline{D}_1(x)$ by

$$\Delta D(x, t) = \overline{D}_2(x, t) - \overline{D}_1(x) \quad (3)$$

The index t denotes the center year of an interval of length p over which the observed data are filtered. Thus

$$\overline{D}_2(x, t) = \sum_{v=-q}^{+q} D(x, t + v)W(v) \quad (4)$$

where $q = (p - 1)/2$. We use $t = 1910, 1911, \dots, 1993$, so that the filtered anomalies cover the 84-year period 1910-1993, while the data used in the filtering extend from 1904-1999. Filter weights are set to zero if data are missing. Since data are not available for 1994-1999, we assigned a missing value code for all post-1993 data. We then stipulate that a filtered mean can only be computed if a critical fraction of the sum of the Gaussian filter weights (W_{crit}) is exceeded at any grid-point over any p -year period. Here $W_{crit} = 0.6$, which allows

us to calculate filtered means for 1993.⁷

The rationale for filtering is to reduce high-frequency variability, e.g., on timescales associated with El Niño behavior and volcanoes. For the purposes of greenhouse-gas detection studies, this is an undesirable noise component. The recent study by Santer et al. (1993) employed overlapping decadal averages as a means of reducing noise. This is equivalent to the use of a square-wave filter. The convolution of observed anomalies and the decadal average filter effectively smoothes out features on timescales of less than 10 years. In the present study, we analyse the significance of signal trends on timescales as short as 10 years, and therefore selected the 13-term Gaussian filter, which preserves more of the variance on timescales of 5-10 years.

The set of time-evolving monthly-mean anomaly fields, $\Delta D(x, t)$, was then used to compute seasonal and annual averages. As in Santer et al. (1993), we stipulated that the grid-point coverage for $\overline{D}_2(x, t)$ must be a subset of the coverage for $\overline{D}_1(x)$ in order to avoid large, spatially non-random increases in coverage from the beginning of the century to the present (see Appendix A). For a given response experiment and a given season, we now have a single pattern characterizing the model temperature signal, $\Delta M(x)$, and a series of time-evolving patterns characterizing observed temperature changes, $\Delta D(x, t)$.

Two different types of measure have been used previously to compare the spatial fields $\Delta M(x)$ and $\Delta D(x, t)$: centered statistics (with the spatial means removed), such as $R(t)$, and uncentered statistics, such as $C(t)$. These are defined by

$$R(t) = \left[\sum_{x=1}^n (\Delta D(x, t) - \widehat{\Delta D}(t)) (\Delta M(x) - \widehat{\Delta M}) \right] / [n s_D(t) s_M] \quad (5)$$

where

$$s_D^2(t) = \sum_{x=1}^n [\Delta D(x, t) - \widehat{\Delta D}(t)]^2 / (n - 1) \quad (6)$$

⁷Data from 1987 through 1999 are used to calculate a filtered mean for 1993. Assuming that data are valid for 1987 through 1993 (inclusive) and missing thereafter, the fraction of weights associated with non-missing data is roughly 62% and thus exceeds W_{crit} . Note that the same value of W_{crit} was used to compute a filtered mean value for $\overline{D}_1(x)$.

is the spatial variance (with s_M^2 defined similarly), and the $\overline{}$ in (5) and (6) indicates a spatial average, and

$$C(t) = \left[\sum_{x=1}^n \Delta D(x,t) \Delta M(x) \right] \left[\sum_{x=1}^n \Delta M(x)^2 \right]^{-1} \quad (7)$$

$R(t)$ is simply a spatial anomaly correlation (similar to statistics used in measuring the ‘skill’ of numerical weather predictions; see, for example, Anderson and van den Dool, 1994) with the statistic centered about the spatial means of the observed and simulated fields. $C(t)$ is the statistic originally defined and used by Barnett and Schlesinger (1987).

$C(t)$ has several desirable properties. First, it is equal to 1.0 when $\Delta D(x,t) = \Delta M(x)$, so that it is a measure of the strength of the model signal in the observed data. Second, unlike $R(t)$, it does not involve a time-varying observed term in the denominator. Thus a trend in $C(t)$ with increasing time can only be due to increasing similarity between the model and observed mean-change fields in the numerator of (7). In contrast, a trend in $R(t)$ with time can be attributable to a change in either the observed mean state and/or the observed spatial variance. Third, $C(t)$, unlike $R(t)$, is not bounded by ± 1.0 . One consequence is that as $R(t)$ approaches ± 1.0 , changes in the amplitude of the observed pattern will not be reflected in linear trends in $R(t)$.

The choice of which type of statistic to use is not clear cut. $C(t)$ has the apparent advantage that it includes information about the mean change field, $\overline{\Delta D}(t)$, which is an important part of the signal in most situations, while $R(t)$ focusses on the pattern of change. We have shown previously (Santer et al., 1993; see also Section 5.1.4) that $C(t)$ can be decomposed into $R(t)$ and $\overline{\Delta D}(t)$ components. In situations where the model-predicted change in global-mean temperature is large relative to the observed changes, the $R(t)$ component of $C(t)$ is much smaller than the $\overline{\Delta D}(t)$ component, so time series of $C(t)$ look very similar to those of $\overline{\Delta D}(t)$ (see, e.g., Santer et al., 1993, Figure 7). In such cases, $C(t)$ cannot be used to address the attribution issue: i.e., if different external forcing mechanisms give rise to similar rates of global mean change, $C(t)$ cannot be used to determine which of the forcings caused the observed change. It is in this situation that $R(t)$, which focusses on the spatial anomalies about the mean state, may provide the information required to discriminate between forcing mechanisms with different pattern signatures but similar rates of global mean change.

In the following, our emphasis is on attribution, and hence we prefer to use $R(t)$. We also compute $C(t)$, since its decomposition provides us with useful information about the relative magnitudes of global-mean change and pattern similarity components for different model-predicted signals from the TP experiments.

5 Comparison of Model and Observed Temperature Change Patterns

In this section, we use both $C(t)$ and $R(t)$ to compare model and observed temperature change patterns. For each of the three TP experiments and for each season, a single pattern characterizing the model signal is compared with 84 observed temperature change patterns consisting of filtered data for the years 1910, 1911, ... 1993. The resulting $C(t)$ and $R(t)$ time series illustrate whether this fixed pattern is increasingly evident in the observed data. Pattern correlations are calculated after excluding grid-points with missing observed data, and with a reference period centered on $t_0 = 1954$. Sensitivity to the choice of reference period is minimal, as shown in Appendix A.

5.1 Pattern Similarity Results

5.1.1 CO₂-only Experiment

Consider first the results for temperature signals from the TP experiment with present-day CO₂ forcing and no anthropogenic sulfate aerosols (Figure 6). If CO₂ forcing were the dominant influence on climate – i.e., if other external forcings and internal natural variability were negligible on timescales appropriate to a slowly-evolving greenhouse warming signal – the $R(t)$ and $C(t)$ time series should show strong multi-decadal positive trends.

$C(t)$ does not show evidence of large, positive trends which are sustained for 40 to 50 years or longer (Figure 6). This is in accord with results obtained by Santer et al. (1993) for the surface temperature signals from CO₂ doubling experiments performed with five different AGCMs. The largest and most sustained $C(t)$ trends occur at the beginning of the record

(between ca. 1910-1940), not at the time of most-rapidly increasing forcing (see Wigley and Raper, 1992). These changes in $C(t)$ parallel those in global mean temperature (Jones et al., 1991; Jones and Briffa, 1992). Further similarities in the behavior of $C(t)$ and global mean temperature are the periods of relatively little change in either between ca. 1940 and 1970 and the increase in both after 1975. Note also that $C(t)$ time series in all seasons except DJF show the effects of the Pinatubo eruption in June 1991 and the volcanic aerosol-induced reduction in global mean temperature (Hansen et al., 1993; Jones, 1994).

These similarities in the behavior of global mean temperature and $C(t)$ are not surprising, and are in agreement with the observed and theoretical results of Santer et al. (1993). As will be shown later (Section 5.1.4), $C(t)$ can be partitioned into two components: an $R(t)$ term and a term related to $\widehat{\Delta D}(t)$, the time-varying observed spatial mean. For the seasonal and annual temperature signals from the C experiment, the $\widehat{\Delta D}(t)$ term is at least several times larger than the $R(t)$ term.

$R(t)$ behaves quite differently from $C(t)$. In all seasons, $R(t)$ increases during the same 1910-40 period over which $C(t)$ rises, but then decreases and shows high-frequency oscillations about some mean state, with no evidence of a large, positive linear trend component over the last 40-50 years. These results are also similar to those presented in Santer et al. (1993). The initial increase in $R(t)$ is in accord with the results of Wigley and Jones (1981) and Kelly et al. (1982), who found that the pattern of observed warming in the 1930s showed some evidence of the high-latitude amplification characteristic of the temperature response in CO₂ doubling experiments.

5.1.2 Sulfate-Only Experiment

The $C(t)$ time series for the sulfate-only temperature signals are virtually the inverse of the $C(t)$ time series for the TP CO₂-only experiment (c.f. Figures 6 and 7). This is not surprising, since we have replaced a spatially-coherent warming pattern by a spatially-coherent cooling pattern in computing the spatial covariance between the fixed model pattern and the time-varying observed fields, and because (apart from the sign) $C(t)$ is expected to parallel $\widehat{\Delta D}(t)$ in both cases.

For $R(t)$, however, the time series are not simply the inverse of the results for the CO₂-

only signals. In JJA and SON, $R(t)$ has a large positive linear trend over 1940-1970, a period over which the CO_2 -only $C(t)$ and $R(t)$ trends are small.

5.1.3 Combined Forcing Experiment

The $R(t)$ time series for the comparison of observed temperature changes and signals from the TP combined forcing experiment are rather different from the corresponding time series for the CO_2 -only experiment (c.f. Figures 6 and 8). This is most pronounced in JJA and SON: $R(t)$ now shows a large positive trend which is sustained over the last 40-60 years of the observed record. For JJA, SON and ANN, $R(t)$ trends are negative over ca. 1910-1940. The positive trends in JJA and SON are similar to those for the sulfate-only case (Figure 7). The implication is that some of the time-increasing congruence between the combined forcing signal pattern and observations is coming from areas where cooling occurs, at least over 1940-1970, possibly related to sulfate aerosol effects.

These results suggest that the inclusion of forcing by both CO_2 and anthropogenic sulfate aerosols enables one to obtain a better fit between observed variations in near-surface temperature and a model-predicted temperature response pattern. This is particularly so over the 1940-1975 period, when $R(t)$ shows little trend for the CO_2 -only signal (Figure 6). In other words, the observed data must have some time-increasing congruence with the time-independent JJA and SON signal patterns from the TP combined forcing experiment.

Some insights into the regions that contribute most to the trends in $R(t)$ are obtained by comparing the linear trends in observed near-surface temperature (e.g., over 1946-86; see Karl et al., 1994) with the TP signal patterns from the combined forcing experiment. For JJA, such a comparison shows that the observed cooling over southern-eastern Europe, the east coast of the U.S., and the Tibetan/Mongolian Plateau is congruent with the TP temperature change pattern (c.f. Figure 2c with Figure 2f of Karl et al., 1994). This must contribute towards the large 50-year trends in $R(t)$ and $C(t)$.

For $C(t)$, the changes for all three perturbation experiments correlate strongly with $\widehat{\Delta D}(t)$ (the observed time-varying spatial mean), although the latter correlations are least for the combined forcing case. $R(t)$ results are only weakly correlated with $\widehat{\Delta D}(t)$ (see Table 4). This points to a dominant influence of $\widehat{\Delta D}(t)$ in the behavior of $C(t)$, as discussed

further below.

5.1.4 Decomposition of $C(t)$

Before addressing the issue of the significance of the linear trends in our pattern correlation statistics (Section 6), we discuss more fully the reasons for $C(t)$ changes. Santer et al. (1993) showed that $C(t)$ could be decomposed into an $R(t)$ term and a term related to $\widehat{\Delta D}(t)$, the time-varying observed spatial mean,

$$C(t) \approx aR(t) + b\widehat{\Delta D}(t) \quad (8)$$

where

$$a = s_D(t)s_M/(s_M^2 + \widehat{\Delta M}^2) \quad (9)$$

and

$$b = \widehat{\Delta M}/(s_M^2 + \widehat{\Delta M}^2) \quad (10)$$

Figure 9 shows the actual values of $C(t)$ for the TP CO₂-only experiment, and the values approximated using equation (8). It is obvious that (8) is an excellent approximation. The ratio b/a may therefore be used as a measure of the relative contributions of the observed spatial-mean change component and the pattern similarity component to the $C(t)$ statistic.

Table 5 indicates that for temperature signals from the experiments with individual CO₂ and sulfate aerosol forcings, b/a is always substantially greater than 1.0, with the largest value (11.98) for annually-averaged data in the CO₂-only experiment. For the CO₂-only and sulfate-only signals, therefore, the pattern similarity component of $C(t)$ is swamped by the change in $\widehat{\Delta D}(t)$, so that $C(t)$ approximately reduces to a record of $\widehat{\Delta D}(t)$ scaled by $1/\widehat{\Delta M}$ (since the s_M^2 term in equation (10) is relatively small – see Table 5).

The ratio b/a is generally much closer to unity for the combined forcing experiment. In

this instance, therefore, $C(t)$ apparently provides some useful information on model versus observed pattern congruence. However, an unambiguous interpretation of trends in $C(t)$ is not possible without decomposition of the statistic, as in equation (8). Even for temperature signals from the SC experiment, a long-term trend in $C(t)$ can be obtained through $\widehat{\Delta D}(t)$ alone – i.e., without a trend in $R(t)$ (see, e.g., DJF results in Figure 8).

6 Significance of $R(t)$ and $C(t)$ trends

In the previous section we showed that for certain seasons, the $R(t)$ and $C(t)$ time series for surface temperature signals from the TP experiment with combined CO_2 and sulfate aerosol forcing show large, positive linear trends over the last 40-60 years. Such trends are qualitatively consistent with the hypothesis that sulfate aerosol-induced cooling could have obscured an enhanced greenhouse effect warming signal. But are these trends statistically significant? This is a difficult question to answer. In order to make meaningful statements about trend significance, we need information about the characteristics of ‘unforced’ $R(t)$ and $C(t)$ trends due solely to the effects of internally-generated natural variability on decadal- to century-timescales. We cannot obtain this information easily from the observed data in view of the difficulties involved in partitioning signal and noise – i.e., separating the observed variability of surface temperatures into a component associated with a time-evolving response to anthropogenic influences and a component associated with natural variability.

In this study, we employ model-generated noise data. In the model world, we can examine output from experiments with no external forcing in order to estimate the magnitude and spatial properties of natural variability noise.

The 20-year TP control integration is clearly too short for estimating natural variability noise characteristics on timescales appropriate to a slowly-evolving anthropogenic signal. Furthermore, it is not clear whether an AGCM coupled to a mixed-layer ocean (which was the experimental configuration used by TP) is capable of realistically simulating the amplitude and spatial structure of internally-generated natural variability on multi-decadal timescales, since such models cannot accurately represent the horizontal and vertical transport of heat, salt and momentum in the global ocean. These processes are likely to strongly influence decadal- to century-timescale variability in the real world.

We therefore used surface temperature results from two recent multi-centennial control integrations performed with fully-coupled atmosphere-ocean GCMs in order to establish the sampling distributions of $R(t)$ and $C(t)$ trends in the absence of external forcing. Near-surface temperature data were taken from a 600-year control integration with the Hamburg A/OGCM,⁸ and a 1,000-year integration of the GFDL coupled model.⁹ These experiments are referred to below as HAMCTL and GFDLCTL, respectively.

A detailed comparison of the near-surface temperature variability in HAMCTL and GFDLCTL is beyond the scope of this paper. However, since data from these integrations are used in the significance testing analysis, we provide an overview of model variability differences in Appendix B, both in global mean terms and in terms of the dominant patterns of variability.

6.1 Determination of Trend Significance

In order to determine the significance of the $R(t)$ and $C(t)$ trends, we need first to establish the sampling distributions of trends in these statistics in the absence of external forcing. The method we use is similar to that employed by Santer et al. (1994a).

We treat HAMCTL and GFDLCTL in the same way that we treated the observed near-surface temperature data, and first define anomalies relative to some reference state of the control run

$$\Delta N(x, t) = \overline{N}_2(x, t) - \overline{N}_1(x) \quad (11)$$

where N denotes noise data from either HAMCTL or GFDLCTL and the reference state $\overline{N}_1(x)$ is the time-average over the entire control run. As in equation (4), data in $\overline{N}_2(x, t)$ were smoothed with a 13-term Gaussian filter, and the index t denotes the center year of an interval of length p ($= 13$). The HAMCTL and GFDLCTL anomalies were then interpolated to the observed data grid, and model data points outside of the observed region for 1954

⁸The model configuration was the ECHAM-1 T21 resolution AGCM coupled to the Large-Scale Geostrophic OGCM.

⁹Here the experimental configuration was the GFDL R15 AGCM coupled to the Bryan-Cox OGCM.

(see Figure A2) were excluded from the analysis.¹⁰ We then substituted $\Delta N(x, t)$, $\widehat{\Delta N}$, etc. for the corresponding observed terms in equations (5-7), and computed pattern correlation time series in the absence of external forcing. These are referred to subsequently as $R_N(t)$ and $C_N(t)$, where the subscript N indicates that the statistics were derived using either HAMCTL or GFDLCTL noise data, and not observed data.

For each pattern correlation statistic, therefore, there is one time series for each season and the annual average, for each of the three TP response experiments, and for each of the two natural variability integrations. This gives a total of 30 $R_N(t)$ time series and 30 $C_N(t)$ time series. Time series utilizing the HAMCTL (GFDLCTL) data are of length 588 (988). As an example, we show the $R_N(t)$ and $C_N(t)$ time series for the TP combined forcing experiment and the HAMCTL natural variability noise (Figure 10). Both sets of time series have considerable variability on 10- to 20-year timescales. When compared with observed changes in $C(t)$ over the last 40- to 50 years ($\approx 0.3 - 0.4$ in the case of the CO₂-only signal; see Figure 6) $C_N(t)$ changes tend to be noticeably smaller (≈ 0.2). For $R(t)$, however, it is more difficult to evaluate whether observed changes are unusual by visual inspection alone.

We next select an array of trend lengths, $L_i = i \times 10$ ($i = 1, \dots, 5$), appropriate to the length of signal trends we wish to evaluate (10-50 years). We then fit linear trends for different L_i to non-overlapping sections of the $R_N(t)$ and $C_N(t)$ time series. The HAMCTL (GFDLCTL) sample sizes therefore range from 58 (98) for 10-year trends to 11 (19) for 50-year trends. For each response experiment, this yields a distribution of slope parameters, $\beta_N(i)$, $i = 1, \dots, 5$, for each statistic, season, and natural variability noise integration. These distributions are the yardsticks that we will use to judge the significance of signal trends, $\beta_S(i)$.

The signal trends are simply the least-squares linear trends for the final 10-50 years of the $R(t)$ and $C(t)$ time series in Figures 6-8 – i.e., the trends over 1984-93, 1974-93, etc. The $R(t)$ time series for the SON signal from the TP combined forcing experiment illustrates this (see Figure 11). While the signal trends over the final 10-20 years of this time series are negative, trends over longer intervals are positive and become increasingly larger.

¹⁰Unlike the situation when dealing with observed data (see Appendix A), there are no coverage changes as a function of time – once the observed data mask for 1954 has been used to exclude model data points from the analysis, this mask does not change with time.

To compute the significance level (p -value) for a selected experiment and trend length L_i , we compare $\beta_S(i)$ with the sampling distribution of $\beta_N(i)$, and determine k_i , the number of times $\beta_N(i) \geq \beta_S(i)$.¹¹ The probability of obtaining the signal trend by chance based on the natural variability manifested in HAMCTL or GFDLCTL is then simply

$$p_i = k_i/m_i \quad (12)$$

where m_i is the sample size for $\beta_N(i)$. Note that the number of independent linear trend samples in HAMCTL and GFDLCTL is likely to be smaller than m_i due to the long decorrelation time of near-surface temperature, both in models and in the observations (see Hegerl et al., 1994).

For the case we are dealing with here – a relatively short natural variability time series and relatively long trend lengths L_i – the use of non-overlapping chunks provides a noisy estimate of the sampling distribution of ‘unforced’ 10- to 50-year linear trends. A much smoother picture is obtained if overlapping chunks are used (as in Wigley and Raper, 1990, and Santer et al., 1994a), although this exacerbates the problem of non-independence of samples. We computed p -values using both non-overlapping chunks and chunks that had the maximum overlap (i.e., by all but one year). The p -values were generally similar in both cases. More importantly, the decisions on the significance of $R(t)$ and $C(t)$ trends did not depend on whether sampling distributions of $\beta_N(i)$ were computed with non-overlapping or overlapping chunks. In the following section, therefore, our discussion is restricted to results obtained using sampling distributions with overlapping chunks.

6.2 Trend Significance Results

CO₂-Only Signal

The p -values for the near-surface temperature signal from the TP CO₂-only experiment are given in Table 6 as a function of trend length, season, statistic, and natural variability noise experiment. Based on noise levels from either of the natural variability experiments,

¹¹Note that this is a one-tailed test, since we have directional information about the signal – the linear trends in $R(t)$ and $C(t)$ should be positive.

linear signal trends for $C(t)$ are significant at the 5% level or better for all seasons and all trend lengths ≥ 20 years. Since we have shown that $C(t)$ basically provides information on observed changes in global-mean temperature for the the CO_2 -only signal (Section 5.1.4), our results imply that the most recent 20- to 50-year trends in observed global-mean temperature are large relative to the variance of ‘unforced’ linear trends in HAMCTL and GFDLCTL.

Numerous other studies have demonstrated that the observed long-term (greater than ≈ 100 years) changes in global-mean, annually-averaged temperature are significant. Such studies have used either statistical models of natural variability (Wigley et al., 1989; Karl et al., 1991; Bloomfield and Nychka, 1992; Allen and Smith, 1994; but see also Woodward and Gray, 1993), or have derived natural variability noise estimates from one-dimensional upwelling-diffusion models (Wigley and Raper, 1990, 1991a,b) or fully-coupled A/OGCMs (Stouffer et al., 1994). Our investigation differs from such previous work in its focus on shorter (10- to 50-year) signal trends, seasonal decomposition of the signal, and the use of noise information from both HAMCTL and GFDLCTL.

The only previous study that has considered recent short-term trends in global-mean annually-averaged temperature is that by Allen et al. (1994), which investigated the significance of the linear trend in global-mean lower tropospheric temperature¹² from 1979-1994. Allen et al. used detrended instrumental sea-surface temperature (SST) data to estimate the magnitude of natural variability on timescales appropriate to the length of their signal, and concluded that the 15-year trend in lower tropospheric temperature was not significant.

Hegerl et al. (1994) used an uncentered pattern similarity statistic to compare observed and model-predicted patterns of temperature trends. As in the present investigation, Hegerl et al. found that their (non-optimized) detection statistic was largely a measure of observed global-mean annually-averaged temperature change for a CO_2 -only signal. The most recent 20- and 30-year trends in the Hegerl et al. detection statistic were highly significant relative to the variance of linear trends in the first 385 years of HAMCTL, which agrees with our findings for 20- to 30-year trends in $C(t)$. Hegerl et al. improved upon their non-optimized results by rotating their fingerprint pattern in a direction in which the GHG signal could be well-represented and the natural variability noise was small.

In contrast, none of our $R(t)$ results for the TP CO_2 -only signal achieve significance at

¹²As sampled by the satellite-based microwave sounding unit (MSU).

the 5% level (Table 6). There is little evidence, therefore, that the spatial pattern (minus the global mean) of temperature-change in the TP integration with CO₂-only forcing is steadily evolving in the observed data.

Combined CO₂/Sulfate Aerosol Signal

The trends in $C(t)$ from the SC experiment are significant in all seasons for trend lengths from 20 to 50 years (Table 7). This result is similar to that obtained in the previous section for the CO₂-only signal, and largely arises through the strong correlation between $C(t)$ and $\widehat{\Delta D}(t)$. However, in the case of the 50-year trend lengths for JJA and SON, it is clear that trends in $R(t)$ also contribute to the $C(t)$ results, since $R(t)$ shows a trend for these two seasons (see Figure 8) and the $R(t)$ component of $C(t)$ is relatively large in SC (see Table 5).

A far more important result is that the $R(t)$ trends for the SC signal are significant for trend lengths of 50 years in JJA and SON. This result does not depend on the model used to define the natural variability noise – i.e., the $\beta_N(i)$ sampling distributions – at least not for the two control runs examined here. It indicates that, in these seasons, there is an evolving expression of the SC signal pattern in the observed data, independent of any trend in global-mean temperature. The 50-year $R(t)$ trend for the annually-averaged SC signal is marginally significant for the GFDLCTL noise ($p = 0.04$) but is not significant relative to the HAMCTL noise ($p = 0.08$). In all seasons, and for both noise integrations, p -values for 50-year $R(t)$ trends are smaller than the p -values for 10- to 40-year trend lengths (see Table 7).

Sulfate-Only Signal

For the linear trends in $R(t)$ from the TP sulfate-only experiment, only one result achieves significance at the 5% level or better: the most recent 50-year trend in SON (see Table 8). Recall that the 50-year $R(t)$ trend in SON also achieves significance for the signal from the combined forcing experiment (Table 7). As noted previously, this suggests that some portion of the multi-decadal $R(t)$ trend in the latter case is due to increasing congruence between areas of predicted and observed cooling.

The $C(t)$ results for the S experiment are generally the opposite of those for the C experiment (c.f. Tables 6 and 8) – i.e., $C(t)$ trends are significantly negative for most trend

lengths and seasons.¹³

7 Summary and Conclusions

Over the past decade, a substantial modelling effort has been directed towards predicting the climate changes likely to result from past and future increases in CO₂ concentrations. An important question is whether the predictions by such models are consistent with the observed surface temperature record over the last century. Can we extract a pattern of change from these observations that is consistent with model predictions for an enhanced greenhouse effect?

The answer to this question depends on how the searched-for GHG signal is defined. Studies that have defined a GHG signal in terms of a pattern that incorporates the global-mean change have showed that this signal is identifiable with a high level of confidence in the observed data (Hegerl et al., 1994). Such investigations have also shown that the global-mean change is the dominant component of the signal, and that it is this component that makes the largest contribution to the overall significance of the results. This complicates the attribution of observed changes to the specific cause of GHG forcing, since different combinations of external forcings and/or natural variability may yield similar global-mean changes.

Other investigations have attempted to address the attribution issue by defining a GHG signal in terms of spatial anomalies about the global-mean change. The rationale here is that it is probably difficult to obtain a high level of correspondence between the small-scale features of observed- and model-predicted changes by mechanisms other than those used in the model forcing experiment. To date, studies that have searched for model-predicted greenhouse warming patterns in observed data (after removal of global-mean changes) have yielded negative or inconclusive results (Santer et al., 1993). One plausible explanation for this result is that the observed record represents the integrated response to a number of external forcing mechanisms (changes in greenhouse gases, solar output, anthropogenic sulfate aerosols, volcanic aerosols, etc.), coupled with internally-generated natural variability.

¹³Differences in the C versus S significance results for $C(t)$ are predominantly due to differences in $\widehat{\Delta M}$, since $C(t) \approx \widehat{\Delta D}(t)/\widehat{\Delta M}$.

The failure of attempts to detect the small-scale spatial structure of a GHG signal may therefore be due to the neglect of other forcings in defining the signal we have been seeking.

Model integrations recently carried out by Taylor and Penner (1994) represent a first attempt to simulate explicitly the temperature response to a combination of CO₂ and sulfate aerosol forcing, and form the basis of the current study. The integrations were performed with the LLNL version of the NCAR CCM1 atmospheric General Circulation Model (AGCM) coupled to a fixed-depth mixed-layer ocean. The AGCM was also coupled to a tropospheric chemistry model (GRANTOUR) which simulated the transport, transformation and removal of various sulfur species. Four integrations were performed: a control run with no anthropogenic sulfate aerosols and (approximately) pre-industrial CO₂ concentrations (CTL), a 'sulfate-only' experiment with near-present-day sulfur emissions (S), a 'CO₂-only' experiment with near-current atmospheric CO₂ concentrations (C), and a combined forcing experiment with near-present-day sulfur emissions and CO₂ levels (SC).

The seasonally- and annually-averaged patterns of near-surface temperature change were very different in the three response experiments. While the integrations with individual forcing showed either global-scale, spatially-coherent cooling (S) or warming (C), the experiment with combined sulfate/CO₂ forcing yielded more complex patterns of temperature change, characterized by both cooling and warming, with most of the cooling restricted to the Northern Hemisphere (see Figures 1-3). Both the S and SC experiments showed relatively weak correlations between the patterns of radiative forcing and near-surface temperature response (Table 1), illustrating the importance of the atmospheric general circulation in modulating the response to regionally-localized forcing. The latter result suggests that detection studies that use the pattern of sulfate aerosol forcing as a proxy for the pattern of near-surface temperature response may yield misleading results.

To determine whether the signal patterns in the S, C and SC experiments were increasingly evident in the observed temperature data, we used an observed data set comprising monthly-mean, land-based surface air temperatures and sea-surface temperatures for the interval 1854-1993 (Jones et al., 1991). Observed temperature fields were expressed as anomalies relative to a 1948-60 reference period and filtered in order to reduce high-frequency noise. Model signals were computed using 20 years of data, and were defined as anomalies relative to a 20-year sample of the CTL experiment. For each experiment and season, a single pattern characterizing the model signal was compared with 84 observed temperature change

patterns for the years 1910, 1911, ... 1993.

Two different forms of pattern similarity statistic were computed: $R(t)$, in which the spatial means of the observed and simulated fields are subtracted, and $C(t)$, which retains these spatial means. The two statistics provide different information in the detection context. $R(t)$, which focusses on spatial anomalies about the global-mean state, can provide the information required to discriminate between forcing mechanisms with different pattern signatures but similar rates of global mean change. $C(t)$ contains both pattern and area-average information. In many cases, the area-average changes dominate the pattern similarity component of $C(t)$, so that it reduces from a multivariate to a univariate statistic. Since the primary goal of this study was attribution of observed climate changes to a specific causes, we therefore focussed on the interpretation of $R(t)$ results.

The time series of $R(t)$ and $C(t)$ indicate whether the model equilibrium signal is becoming increasingly evident in the observed data. For the CO₂-only signal, none of the $R(t)$ time series show evidence of a positive trend sustained over at least 4-5 decades. In contrast, the $R(t)$ time series for the SC signal showed large, multi-decadal positive trends in JJA and SON. This indicates that in these seasons there is an evolving expression of the SC signal pattern in the observed data, independent of any trend in global-mean temperature. 50-year trends are smaller than in the case of the CO₂-only signal. For the sulfate-only experiment, $R(t)$ has a large increase over 1940-70 in JJA, SON, and ANN, suggesting that at least some portion of the long-term trends in $R(t)$ in JJA and SON for the SC signal is attributable to areas where cooling occurs.

In virtually all cases (the exceptions are the JJA and SON signals from the SC experiment), the behavior of $C(t)$ closely parallels changes in observed global-mean temperature. $C(t)$ results are in accord with the findings of Santer et al. (1993).

We tested the sensitivity of these results to different choices of t_0 , the central year of the reference period used for defining observed near-surface temperature changes. While the absolute values of $C(t)$ and $R(t)$ depend on the choice of t_0 , the changes in both statistics as a function of time are to first order independent of t_0 .

In the absence of reliable information on the magnitude and spatial characteristics of long timescale natural variability in the real world, we used data from multi-century control integrations performed with the Hamburg and GFDL fully-coupled A/OGCMs ('HAMCTL';

Cubasch et al., 1992; Hegerl et al., 1994; 'GFDLCTL'; Delworth et al., 1993; Stouffer et al., 1994) to assess the significance of the most recent $R(t)$ and $C(t)$ trends (β_S) for the signals from the S, C, and SC experiments. HAMCTL and GFDLCTL provide internally-consistent (but model-specific) estimates of the magnitude and patterns of surface temperature variability on long timescales. In the first stage of the significance analysis, we correlated 588 (988) years of filtered near-surface temperature anomaly patterns from HAMCTL (GFDLCTL) with the signals from the three TP perturbation experiments. By fitting linear trends to discrete sections of the resulting $R_N(t)$ and $C_N(t)$ time series, we next obtained sampling distributions of β_N (the pattern correlation statistic trends in the absence of external forcing). The significance of the β_S signal trends was then determined by comparison with the appropriate β_N sampling distribution.

For the temperature-change signal from the TP combined forcing experiment, our results indicated that the most recent 50-year $R(t)$ trends are significant in JJA and SON. This result did not depend on whether HAMCTL or GFDLCTL was used to estimate β_N . For the signal from the TP sulfate-only experiment, the 50-year $R(t)$ trend in SON is the only result which achieves significance at the 5% level or better. None of the $R(t)$ trends for the CO₂-only signal was significant.

This analysis supports but does not prove that we have detected, beyond doubt, an anthropogenic climate change signal in observed records of near-surface air temperature change. We stress that we are relying on model noise and model signals in order to assess trend significance. Both have attendant uncertainties (Santer et al., 1994a). Here, the natural variability simulated in the HAMCTL and GFDLCTL integrations constituted the yardstick used for judging whether a signal trend was unusual. Some of the long-term variability in both experiments may be related to residual drift rather than *bona fide* internally-generated fluctuations of the coupled system (see Figure B1; also Cubasch et al., 1994; Santer et al., 1994a).¹⁴ Since the critical multi-decadal to century-timescale natural variability in HAMCTL and GFDLCTL has not been validated (and may be difficult to validate, given our poor knowledge of the spatial characteristics of natural variability on decadal- to century-timescales), one must be cautious not to overinterpret the significance levels obtained here.

¹⁴If the drift behavior over the first 200 years of HAMCTL inflates values of β_N , our significance estimates may well be conservative, at least for $C(t)$. It is more difficult to know how drift behavior affects significance estimates for $R(t)$.

Furthermore, although we have tried to look at the model-dependence of natural variability noise by considering results from both HAMCTL and GFDLCTL, we have little or no information about the sensitivity of the variability to changes in resolution, physics, flux correction scheme, or parameterizations. A preliminary comparison (see Appendix B) suggests that the low-frequency variability of surface temperature is very different in HAMCTL and GFDLCTL, both in global mean terms and in terms of the dominant patterns of near-surface temperature variability. These differences in variability do not affect the significance of the 50-year $R(t)$ trends in JJA and SON for the temperature signal from the SC experiment. Both integrations, however, employed large flux corrections to compensate for systematic errors in their atmospheric and oceanic components (Gates et al., 1993). It is not inconceivable that such flux adjustment procedures can have an impact on the simulated multi-decadal-to century-timescale variability.

Similar uncertainties affect the signal patterns used here. As has been noted by Taylor and Penner (1994), the predicted pattern of temperature change in the SC experiment is clearly a function of the relative magnitudes of the positive forcing by greenhouse gases and the negative forcing due to anthropogenic sulfate aerosols. While the present-day CO₂ forcing is relatively well-known, the current sulfate aerosol forcing is uncertain by a factor of at least two (Wigley and Raper, 1992; Penner et al., 1994b). Furthermore, relative to other studies (Charlson et al., 1991; Kiehl and Briegleb, 1993), the aerosol forcing in TP is quite large, and is at the upper end of the range used by Wigley and Raper (1992). Clearly, these forcing uncertainties translate into considerable uncertainty regarding the spatial pattern of the temperature response in the SC experiment, and in the relative contributions to this response from CO₂ and aerosols. Further signal uncertainties include the lack of a dynamic ocean, the neglect of the indirect effects of sulfate aerosols, the omission of the radiative effects of trace gases other than CO₂ (which may have temperature-change signatures that differ from that for CO₂ alone; see Wang et al., 1991), and failure to include the effects of carbonaceous aerosols generated by biomass burning, fossil fuel combustion, and industrial processes (Penner et al., 1992, 1994b).¹⁵

We also find that substantial natural variability can occur even in the experimental configuration used by TP: an AGCM coupled to a fixed-depth mixed-layer ocean with prescribed meridional heat transport. This “noise” (i.e., statistical signal uncertainty) is superimposed

¹⁵The optical properties of sulfate aerosols are very different from those of carbonaceous aerosols.

on the signals used here, and its amplitude is not negligible relative to the amplitude of the temperature-change signal in the SC experiment. If the variability of near-surface temperature is similar to that generated in much longer integrations by comparable AGCM/mixed-layer ocean configurations (see Hansen et al., 1988, 1994), an integration time of 20 years or less may be insufficient to obtain a reliable estimate of the true, underlying SC signal.

Future work should use longer integrations to address the robustness of the signal patterns employed here, and should consider whether the signal-to-noise ratios obtained in our current work can be enhanced by the application of optimal detection techniques (Hasselmann, 1979, 1993), which attempt to filter out those signal components that are contaminated by natural variability noise.

The caveats regarding the signals and natural variability noise which form the basis of this study are numerous. Nevertheless, we have provided first evidence that both the largest-scale (global-mean) and smaller-scale (spatial anomalies about the global mean) components of a combined CO₂ / anthropogenic sulfate aerosol signal are identifiable in the observed near-surface air temperature data. If the coupled-model noise estimates used here are realistic – an assumption which thus far has not been rigorously tested on the crucial decadal- to century-timescales – we can be confident that the anthropogenic signal which we have identified is distinctly different from natural variability noise. The fact that we have been able to detect the detailed spatial signature in response to combined CO₂ and sulfate aerosol forcing, but not in response to CO₂ forcing alone, suggests that some of the regional-scale background noise (against which we were trying to detect a CO₂-only signal) is in fact part of the signal of a sulfate aerosol effect on climate.

While our confidence in the identification of an anthropogenic effect on climate is high, we have not shown conclusively that the signal identified can be attributed to the unique cause of anthropogenic sulfate aerosols and CO₂. We have taken a first step in the direction of attribution by showing that agreement between modelled and observed changes exists at relatively small spatial scales. Enhancing our confidence in the attribution of observed climate changes to anthropogenic causes will require the elimination of other forcing mechanisms (such as solar variability and volcanoes) as possible explanations for the observed changes. This should provide strong motivation for future modelling studies.

Acknowledgments

Work at Lawrence Livermore National Laboratory was performed under the auspices of the U.S. Dept. of Energy, Environmental Sciences Division, under contract W-7405-ENG-48. Support for JEP was provided by the CHAMMP program; TMLW and PDJ were supported by the U.S. Dept. of Energy, Environmental Sciences Division, under grant no. DE-FG02-86ER60397. The EPRI MECCA Program supplied the computer time necessary to extend the TP integrations. The authors are grateful to Klaus Hasselmann, Wolfgang Brüggemann, and Gabi Hegerl (Max-Planck Institute for Meteorology) for their comments and helpful suggestions, and to Eduardo Zorita and Jin-Song Xu for supplying results from the 600-year HAMCTL integration. Ron Stouffer and Tom Delworth kindly supplied data from the 1,000-year GFDLCTL integration. All color graphics were produced with software developed at the Program for Climate Model Diagnosis and Intercomparison by Dean Williams and Bob Mobley.

Appendix A: Sensitivity to Choice of Reference Period

It is useful to consider how the choice of t_0 – the central year of the p -year period used for computing the reference state $\overline{D}_1(x)$ – influences the $R(t)$ and $C(t)$ results for the near-surface temperature signals from the three TP response experiments. Sensitivity of trends in $R(t)$ and $C(t)$ to the choice of reference state would make interpretation of the results more difficult.

In the ideal situation, in which there are no missing values in the observed data set, it can be shown analytically that the changes in $C(t)$ as a function of time are independent of the choice of t_0 . The same holds true for $R(t)$ if the observed spatial variance, $s_D^2(t)$, is roughly constant with time (as it is for temperature data; see Santer et al., 1993).

However, the observed data do not comply with this ideal scenario, and show large, spatially non-random changes in coverage as a function of time. This is illustrated in Figures A1 and A2. Since the maximum spatial coverage of observed data is determined by the coverage over the reference period,¹⁶ changes in coverage with time will mean that the data

¹⁶This is because the subset of data in $\overline{D}_2(x, t)$ that is actually used must be equal to or a subset of the

employed in calculating $C(t)$ and $R(t)$ will depend on the choice of t_0 , so results may be sensitive to t_0 .

Furthermore, it may be difficult to determine whether a positive trend in $C(t)$ or $R(t)$ reflects a true time-increasing pattern similarity between the model and observed fields or the effect of coverage changes. In choosing t_0 , it is therefore desirable to minimize coverage changes with time, and yet still maintain adequate coverage.

The effect of different choices of t_0 and attendant differences in coverage is illustrated in Figures A3 ($R(t)$) and A4 ($C(t)$), by considering the extreme cases $t_0 = 1906$ and $t_0 = 1954$. Results are for the seasonally- and annually-averaged near-surface temperature signals in the TP experiment with combined CO_2 /sulfate aerosol forcing. For both statistics, it is clear that the changes in model-versus-observed spatial covariance as a function of time (i.e., $\partial R/\partial t$ and $\partial C/\partial t$) are, to first order, independent of t_0 .

These results have several important implications. First, it is apparent that linear trends in the pattern correlation time series are relatively insensitive to differences in observed data coverage for the two choices of t_0 . Second, we note that the absolute values of both statistics are somewhat arbitrary, and can be raised or lowered by selection of a different reference state. This is not a problem here, since our interest is in trends rather than absolute values.

Since our focus is on trends over the last 10- to 50 years of the pattern correlation time series, we use $t_0 = 1954$. This guarantees that we have relatively stable observed data coverage over 1944-1993, and considerably higher coverage than for $t_0 = 1906$ (see Figures A1 and A2). The choice of a later t_0 would improve coverage still further, but have the undesirable effect of yielding large coverage changes in the 1950s (due to the introduction of data from Antarctica).

coverage of the reference state $\overline{D}_1(x)$ in order to calculate changes relative to $\overline{D}_1(x)$.

Appendix B: Comparison of Variability in HAMCTL and GFDLCTL

Features of the variability of near-surface air temperature in the first 100-385 years of HAMCTL have been described by Cubasch et al. (1992, 1994), Santer et al., (1994a,b) and Hegerl et al. (1994). HAMCTL exhibits substantial variability on decadal-to-century time scales (see Figure B1a). However, part of this variability, particularly over the first 200 years of the integration, may be attributable to climate drift rather than *bona fide* variability of the coupled ocean-atmosphere system.

The variability in GFDLCTL (Figure B1b) has been analysed by Delworth et al. (1993) for the thermohaline circulation and by Stouffer et al. (1994), Mehta and Delworth (1994), and Manabe and Stouffer (1995) for near-surface air temperature and SST. These investigations have attempted to validate the interannual- to interdecadal timescale variability simulated by the GFDL A/OGCM. The study by Stouffer et al. (1994) showed qualitative agreement between observed and model-simulated patterns of interannual near-surface temperature variability,¹⁷ and also indicated that model and observed relationships between global-mean and regional temperature changes show a broadly similar structure. Mehta and Delworth (1994) found considerable similarity between simulated and observed patterns characterizing decadal-timescale variability of SST patterns in the tropical Atlantic. Such studies help to build confidence in the reliability of the model variability on interannual to decadal timescales. Validation of model variability on longer timescales is more problematical.

The behavior of global-mean, annually-averaged near-surface temperature is non-stationary in both HAMCTL and GFDLCTL, despite the application of flux correction schemes. The linear trend in GFDLCTL ($0.023^{\circ}\text{C}/\text{century}$) is relatively monotonic, while the much larger trend in HAMCTL ($0.102^{\circ}\text{C}/\text{century}$), is primarily due to a large negative excursion of temperature over the first two hundred years of the integration. The standard deviations of temperature for the detrended data differ by only 0.02°C (0.10°C for GFDLCTL versus 0.12°C for HAMCTL). For the observations over the period 1861-1993, the corresponding figure is 0.21°C if no overall trend is removed. If an EBM is used to remove a combined GHG/sulfate aerosol effect from the observed data in an optimal way (c.f. Wigley and

¹⁷With the exception of the tropical Pacific, where both GFDLCTL and HAMCTL underestimate the variability associated with ENSO phenomena.

Raper, 1991b), the standard deviation of the residuals is 0.10°C, a result that is robust to uncertainties in the assumed forcing history.

Power spectra of the detrended global-mean, annually-averaged near-surface temperature changes for HAMCTL and GFDLCTL indicate that GFDLCTL has more power than HAMCTL on timescales of roughly 2-60 years, and less power than HAMCTL for periods greater than ≈ 60 years (see Figure B2).¹⁸ Part of the explanation for the difference in power at the low-frequency end of the coupled model spectra is the large negative temperature excursion over the first 200 years of HAMCTL.¹⁹ The areas defined by the upper and lower 95% confidence limits for each of the coupled model spectra overlap at all frequencies, with minimum overlap at the high- and low-frequency ends of the spectra.

Figure B2 also shows the spectrum from a 5,000-year integration of an upwelling-diffusive energy-balance model (EBM; Wigley and Raper, 1990; 1992) with random forcing and a climate sensitivity, ΔT_{2xCO_2} , of 3.0°C for a doubling of atmospheric CO₂. As noted by Hasselmann (1976) and Wigley and Raper (1990), the low frequency spectrum depends on ΔT_{2xCO_2} . The GFDL AGCM (coupled with a mixed-layer ocean) has a sensitivity of 3.5°C (Stouffer, personal communication), while the corresponding figure is 2.6°C for the ECHAM AGCM (see Table B1 in Gates et al., 1992). The EBM value was chosen to fall roughly between the Hamburg and GFDL sensitivities.²⁰ Note that in the lower frequency range, HAMCTL has higher power than GFDLCTL despite having a lower value of ΔT_{2xCO_2} - a result opposite to that expected on theoretical grounds, as the EBM result on the left-hand side of Figure B2 indicates. This is at least partially due to the above-mentioned climate drift in the first 200 years of HAMCTL.

In contrast to the HAMCTL and EBM results, the spectrum for GFDLCTL flattens out at periods greater than 100 years. The ubiquity of substantial power at century timescales in paleoclimate spectra (see, e.g., Crowley and North, 1991; Stocker and Mysak, 1992) suggests that either the paleodata have some low-frequency forcing that the model is lacking (e.g., solar variability, volcanoes) or that the model is underestimating the magnitude of century-

¹⁸These results are relatively insensitive to the choice of window width for smoothing.

¹⁹Removal of the overall linear trend reduces, but does not remove the effect of this initial drift on the estimated spectrum.

²⁰At high frequencies, the EBM was tuned to agree with the observed data after factoring out the effect of ENSO (which accounts for about 30% of the variance at timescales less than 10 years).

timescale variability associated with internal oscillations of the coupled atmosphere-ocean system.²¹ Note that the EBM also shows a flattening of the spectrum, but at frequencies substantially lower than in the GFDL model simulation (and beyond the lowest frequency shown in Figure B2).

Since we have shown that $C(t)$ closely parallels the ratio between model and observed global-mean temperature changes, and since the standard error for a linear trend of length L_i is related to the spectral power at frequency $1/L_i$, one might expect the significance of $C(t)$ values in Tables 6-8 to differ noticeably depending on whether HAMCTL or GFDLCTL was used to estimate the noise. This is difficult to determine on the basis of p -values, since these are generally highly significant for observed $C(t)$ trends (Tables 6-8). We therefore computed signal-to-noise ratios for $C(t)$ trends of different length. As in Santer et al. (1994a), we define the signal-to-noise ratio SN_i for trend length $i \times 10$ as

$$SN_i = \beta_S(i) / \varepsilon_N(i) \quad (13)$$

where $\beta_S(i)$ is the observed trend in $C(t)$ and $\varepsilon_N(i)$ is the standard error of the sampling distribution of $\beta_N(i)$, computed using either the GFDLCTL or HAMCTL data (see Section 6.1 for full details). The standard error is simply

$$\varepsilon_N(i) = \left[(m_i - 1)^{-1} \sum_{c=1}^{m_i} \beta_N(i, c)^2 \right]^{1/2} \quad (14)$$

where m_i is the sample size (the number of linear trends of length $i \times 10$ in the $C_N(t)$ natural variability time series).

Figure B3 shows SN_i for 10- to 50-year linear trends in $C(t)$. Results are for annually-averaged signal and noise data. For temperature-change signals from the sulfate-only and CO₂-only experiments, the higher GFDL variance on 10- to 50-year timescales (Figure B2)

²¹Errors in the paleodata provide a further possible explanation for this discrepancy. No proxy-based reconstruction can perfectly retrieve the original climatic information, and statistical manipulation of the paleodata may alter their low-frequency characteristics (although this is more likely to have decreased low-frequency power than to have amplified it). An additional problem is that paleospectra are for individual sites, not for global-mean temperature.

yields consistently lower signal-to-noise ratios than for ECHAM. For trends ≥ 20 years, SN_i is much larger than the 5% significance threshold (for an assumed Gaussian distribution), so that GFDL versus ECHAM variance differences do not influence decisions on the significance of observed $C(t)$ trends. As discussed previously, $C(t)$ provides a mix of pattern- and global-mean change information in the case of the SC experiment. In this instance, therefore, differences in SN_i cannot be interpreted solely in terms of the different behavior of GFDL and ECHAM global-mean temperature spectra in Figure B2.

Signal-to-noise ratios for 10- to 50-year linear trends in $R(t)$ are surprisingly similar, at least for JJA (Figure B4).²² This suggests that the patterns of near-surface temperature variability on 10- to 50-year timescales must show some degree of correspondence in the GFDL and ECHAM control runs. To compare the dominant patterns of variability, we computed Empirical Orthogonal Functions (EOFs) of the HAMCTL and GFDLCTL annually-averaged near-surface temperature data.²³ Model anomaly data were filtered as described in Section 4, and grid-points outside the observed data mask for 1954 were excluded (see Figure A2). This procedure allows us to compare model variability modes over the same data window used for computing observed $R(t)$ trends.

The partitioning of total space-time variance is similar in both integrations.²⁴ EOFs 1 and 2 of GFDLCTL explain 13.2% and 8.9% of the variance (c.f. 11.0% and 9.9% for HAMCTL). The total number of EOFs required to explain $\geq 95\%$ of the variance is 69 for GFDLCTL and 63 for HAMCTL. The EOF patterns show some large-scale spatial similarity: both coupled models have large loadings of the same sign over high-latitude land areas of North America and Eurasia. There are also pronounced differences. EOF 1 of GFDLCTL has the same sign at virtually all grid-points, while such global-scale coherence is absent in the HAMCTL EOFs. The marked dipole structure over North America in the first pair of HAMCTL EOFs has no analogue in the dominant GFDL modes. Further work is required to identify 'common' patterns of variability in the two integrations (e.g., Common Principal Components; Sengupta and Boyle, 1993) and their characteristic timescales, and to deter-

²²Note that the p -values for $R(t)$ trends (Tables 6-8) are also similar for the GFDL and ECHAM noise estimates.

²³The first 400 years of GFDLCTL and the last 400 years of HAMCTL were used for computing EOFs

²⁴A very different result is obtained if the full global fields are used for computing EOFs. In this case differences in the behavior of sea-ice in the two integrations yield a much flatter eigenvalue spectrum for GFDL than for ECHAM.

mine why the GFDL and ECHAM signal-to-noise ratios (and p -values) are similar for trends in $R(t)$.

References

- Allen MR, Smith LA (1994) Investigating the origins and significance of low-frequency modes of climate variability. *Geophys Res Lett* 21: 883-886
- Allen MR, Mutlow CT, Blumberg GMC, Christy JR, McNider RT, Llewellyn-Jones DT (1994) Global change detection. *Nature* 370: 24-25
- Anderson JL, van den Dool HM (1994) Skill and return of skill in dynamic extended-range forecasts. *Mon Wea Rev* 122: 507-516
- Barnett TP (1986) Detection of changes in global tropospheric temperature field induced by greenhouse gases. *J Geophys Res* 91: 6659-6667
- Barnett TP, Schlesinger ME (1987) Detecting changes in global climate induced by greenhouse gases. *J Geophys Res* 92: 14772-14780
- Benkovitz C (1982) *Compilation of an inventory of anthropogenic emissions in the United States and Canad.* *Atmos Environ* 16: 1551-1563
- Bloomfield P, Nychka D (1992) Climate spectra and detecting climate change. *Climatic Change* 21: 275-288
- Charlson RJ, Langer J, Rodhe H, Leovy CB, Warren SG (1991) Perturbation of the Northern Hemisphere radiative balance by backscattering from anthropogenic sulfate aerosols. *Tellus* 43: 152-163
- Charlson RJ, Schwartz SE, Hales JM, Cess RD, Coakley JA, Hansen JE, Hofmann DJ (1992) Climate forcing by anthropogenic aerosols. *Science* 255: 423-430
- Charlson RJ, Wigley TML (1994) Sulfate aerosol and climate change. *Sci Amer* 270(2): 48-57
- Crowley TJ, North GR (1991) *Paleoclimatology.* Oxford University Press, New York, 339 pp

- Cubasch U, Hasselmann K, Höck H, Maier-Reimer E, Mikolajewicz U, Santer BD, Sausen R (1992) Time-dependent greenhouse warming computations with a coupled ocean-atmosphere model. *Clim Dyn* 8: 55-69
- Cubasch U, Santer BD, Hellbach A, Hegerl G, Höck H, Maier-Reimer E, Mikolajewicz U, Stössel A, Voss R (1994) Monte Carlo climate change forecasts with a global coupled ocean-atmosphere model. *Clim Dyn* 10: 1-19
- Delworth T, Manabe S, Stouffer RJ (1993) Interdecadal variability of the thermohaline circulation in a coupled ocean-atmosphere model. *J Climate* 6: 1993-2011
- Engardt M, Rodhe H (1993) A comparison between patterns of temperature trends and sulfate aerosol pollution. *Geophys Res Lett* 20: 117-120
- Fichefet T, Tricot C (1992) Influence of starting date of model integration on projections of greenhouse-gas-induced climatic change. *Geophys Res Lett* 19: 1771-1774
- Folland, CK, Karl TR, Nicholls N, Nyenzi BS, Parker DE, Vinnikov K Ya (1992) Observed Climate variability and change. In: Houghton JT, Callander BA, Varney SK (eds). *Climate Change 1992. The supplementary report to the IPCC scientific assessment*. Cambridge University Press, Cambridge, 135-170
- Gates WL, Cubasch U, Meehl GA, Mitchell JFB, Stouffer RJ (1993) An intercomparison of selected features of the control climates simulated by coupled ocean-atmosphere general circulation models. *World Climate Research Programme, WCRP-82, Geneva, Switzerland, 46 pp*
- Gates WL, Mitchell JFB, Boer GJ, Cubasch U, Meleshko VP (1992) Climate modelling, climate prediction and model validation. In: Houghton JT, Callander BA, Varney SK (eds). *Climate Change 1992. The supplementary report to the IPCC scientific assessment*. Cambridge University Press, Cambridge, 97-134
- Hansen J, Sato M, Ruedy R (1994) Long-term changes of the diurnal temperature cycle: Implications about mechanisms of global climate change. *Atmos Res (Submitted)*
- Hansen J, Lacis A, Ruedy R, Sato M, Wilson H (1993) How sensitive is the world's climate? *Nat Geog Res Explor* 9: 142-158

- Hansen J, Fung I, Lacis A, Rind D, Lebedeff S, Ruedy R, Russell G, Stone, P (1988) Global climate changes as forecast by Goddard Institute for Space Studies three-dimensional model. *J Geophys Res* 93: 9341-9364
- Hansen J, Lacis A, Rind D, Russel G, Stone P, Fung I, Ruedy R, Lerner J (1984) Climate sensitivity: Analysis of feedback mechanisms. In: Hansen J, Takahasi T (eds) *Climate Processes and Climate Sensitivity*. Maurice Ewing Series 5, American Geophysical Union, Washington D.C., pp 130-163
- Hasselmann K (1993) Optimal fingerprints for the detection of time dependent climate change. *J Climate* 6: 1957-1971
- Hasselmann K (1979) On the signal-to-noise problem in atmospheric response studies. In: Shaw DB (ed) *Meteorology of Tropical Oceans*. Roy Met Soc London, pp 251-259
- Hasselmann K (1976) Stochastic climate models. Part I: Theory. *Tellus* 28:473-485
- Hasselmann K, Sausen R, Maier-Reimer E, Voss R (1993) On the cold start problem in transient simulations with coupled ocean-atmosphere models. *Clim Dyn* 9: 53-61
- Hegerl GC, Storch Hv, Hasselmann K, Santer BD, Cubasch U, Jones PD (1994) Detecting anthropogenic climate change with an optimal fingerprint method. *J Climate* (Submitted)
- Hunter DE, Schwartz SE, Wagoner R, Benkovitz CM (1993) Seasonal, latitudinal, and secular variations in temperature trend: Evidence for influence of anthropogenic sulfate. *Geophys Res Lett* 20: 2455-2458
- Jenkins GM, Watts DG (1968) *Spectral analysis and its applications*. Holden-Day, San Francisco, 525 pp
- Jones PD (1994) Recent warming in global temperature series. *Geophys Res Lett* 21: 1149-1152
- Jones PD, Briffa, KR (1992) Global surface air temperature variations during the twentieth century: Part 1, spatial, temporal and seasonal details. *The Holocene* 2: 165-179

- Jones PD, Wigley TML, Farmer G (1991) Changes in hemispheric and regional temperatures 1851 to 1988 over land and marine areas. In: Schlesinger ME (ed) Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations. Elsevier, Amsterdam, 153-172
- Karl TR, Knight RW, Kukla G, Gavin J (1994) Evidence for radiative effects of anthropogenic sulfate aerosols in the observed climate record. In: Charlson R, Heintzenberg J (eds) Aerosol Forcing of Climate. John Wiley and Sons (In press)
- Karl TR, Heim RR Jr, Quayle RG (1991) The greenhouse effect in Central North America: If not now, when? *Science* 251: 1058-1061
- Kelly PM, Jones PD, Sear CB, Cherry BSG, Tavakol RK (1982) Variations in surface air temperatures, Part 2: Arctic regions, 1881-1980. *Mon Wea Rev* 110: 71-83
- Kiehl JT, Briegleb BP (1993) The relative role of sulfate aerosols and greenhouse gases in climate forcing. *Science* 260: 311-314
- MacCracken MC, Moses H (1982) The first detection of carbon dioxide effects: Workshop Summary, 8-10 June 1981, Harpers Ferry, West Virginia. *Bull Amer Met Soc* 63:1164-1178
- Madden RA, Ramanathan V (1980) Detecting climate change due to increasing carbon dioxide. *Science* 209:763-768
- Manabe S, Stouffer RJ (1995) Low-frequency variation of surface air temperature in a 1,000-year integration of a coupled ocean-atmosphere model. (In preparation)
- Manabe S, Stouffer RJ (1980) Sensitivity of a global climate model to an increase in the CO₂ concentration in the atmosphere. *J Geophys Res* 85: 5529-5554
- Meehl GA, Washington WM, Karl TR (1993) Low-frequency variability and CO₂ transient climate change. *Clim Dyn* 8: 117-133
- Mehta VM, Delworth T (1994) Decadal variability of the tropical Atlantic Ocean surface temperature in shipboard measurements and in a global ocean-atmosphere model. *J. Climate* (Accepted)
- Mitchell JFB, Davis RA, Ingram WJ, Senior CA (1994) On surface temperature, greenhouse gases and aerosols: Models and observations (In preparation).

- Parker DE, Jones PD, Folland CK, Bevan A (1994) Interdecadal changes of surface temperature since the late nineteenth century. *J Geophys Res* 99: 14373-14399
- Penner JE, Dickinson R, O'Neill C (1992) Effects of aerosol from biomass burning on the global radiation budget. *Science* 256, 1432-1434
- Penner JE, Atherton CA, Graedel TE (1994a) Global emissions and models of photochemically active compounds. In: Prinn R (ed) *Global Atmospheric-Biospheric Chemistry*, Plenum Publishing, New York, pp 223-248
- Penner JE, Charlson RJ, Hales JM, Laulainen NS, Leifer R, Novakov T, Ogren J, Radke LF, Schwartz SE, Travis L (1994b) Quantifying and minimizing uncertainty of climate forcing by anthropogenic aerosols. *Bull Am Met Soc* 75: 375-400
- Preisendorfer RW, Barnett TP (1983) Numerical model-reality intercomparison tests using small-sample statistics. *J Atmos Sci* 40: 1884-1896
- Roeckner E, Siebert T, Feichter J (1994) Climatic response to anthropogenic sulfate forcing simulated with a general circulation model. In: Charlson R, Heintzenberg J (eds) *Aerosol Forcing of Climate*. John Wiley and Sons (In press)
- Santer BD, Wigley TML (1990) Regional validation of means, variances and spatial patterns in General Circulation Model control runs. *J Geophys Res* 95: 829-850
- Santer BD, Wigley TML, Jones PD (1993) Correlation methods in fingerprint detection studies. *Clim Dyn* 8: 265-276
- Santer BD, Mikolajewicz U, Brüggemann W, Cubasch U, Hasselmann K, Höck H, Maier-Reimer E, Wigley TML (1994a) Estimates of detection period and detection time for ocean greenhouse warming signals. *Journal of Geophysical Research* (Accepted)
- Santer BD, Brüggemann W, Cubasch U, Hasselmann K, Höck H, Maier-Reimer E, Mikolajewicz U (1994b) Signal-to-noise analysis of time-dependent greenhouse warming experiments. Part 1: Pattern analysis. *Clim Dyn* 9: 267-285
- Schimel DS, Wigley TML, Heimann M, Raynaud D, Alves D, Siegenthaler U (1994) The carbon cycle. In: Houghton JT, Callander BA (eds) *Radiative Forcing of Climate Change: IPCC 1994 WGI Interim Report*. Cambridge University Press, Cambridge (In preparation)

- Schlesinger ME, Mitchell JFB (1987) Climate model simulations of the equilibrium climatic response to increased carbon dioxide. *Rev of Geophys* 25: 760-798
- Sengupta SK, Boyle JS (1993) Statistical intercomparison of global climate models: A common principal component approach. PCMDI Report No. 13, Lawrence Livermore National Laboratory, Livermore, California, 41 pp
- Spiro PA, Jacob DJ, Logan JA (1992) Global inventory of sulfur emissions in the United States and Canada. *J Geophys Res* 97: 6023-6036
- Stocker TF, Mysak LA (1992) Climatic fluctuations on the century time scale: A review of high-resolution proxy data and possible mechanisms. *Clim Change* 20: 227-250
- Stouffer RJ, Manabe S, Vinnikov K Ya (1994) Model assessment of the role of natural variability in recent global warming. *Nature* 367: 634-636
- Stouffer RJ, Manabe S, Bryan K (1989) Interhemispheric asymmetry in climate response to a gradual increase of atmospheric CO₂. *Nature* 342: 660-662
- Taylor KE, Ghan SJ (1992) An analysis of cloud liquid water feedback and global climate sensitivity in a general circulation model. *J Climate* 5: 907-919
- Taylor KE, Penner JE (1994) Response of the climate system to atmospheric aerosols and greenhouse gases. *Nature* 369: 734-737
- Walton JJ, MacCracken MC, Ghan SJ (1988) A global-scale Lagrangian trace species model of transport, transformation and removal processes. *J Geophys Res* 93: 8339-8354
- Wang W-C, Dudek MP, Liang X-Z, Kiehl JT (1991) Inadequacy of effective CO₂ as a proxy in simulating the greenhouse effect of other radiatively active gases. *Nature* 350: 573-577
- Washington WM, Meehl GA (1989) Seasonal cycle experiments on the climate sensitivity due to a doubling of CO₂ with an atmospheric general circulation model coupled to a simple mixed layer ocean model. *J Geophys Res* 89: 9475-9503
- Wigley TML (1991) Could reducing fossil-fuel emissions cause global warming? *Nature* 349: 503-506.
- Wigley TML (1989) Possible climate change due to SO₂-derived cloud condensation nuclei. *Nature* 339: 365-367

- Wigley TML, Jones PD (1981) Detecting CO₂-induced climatic change. *Nature* 292: 205-208
- Wigley TML, Raper, SCB (1990) Natural variability of the climate system and detection of the greenhouse effect. *Nature* 344: 324-327
- Wigley TML, Raper, SCB (1991a) Internally-generated natural variability of global-mean temperatures. In Schlesinger ME (ed). *Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations*. Elsevier, Amsterdam, 471-482
- Wigley TML, Raper, SCB (1991b) Detection of the enhanced greenhouse effect on climate. *Proceedings of the Second World Climate Conference, World Meteorological Organization, Geneva.*
- Wigley TML, Raper, SCB (1992) Implications for climate and sea level of revised IPCC emissions scenarios. *Nature* 357: 293-300
- Wigley TML, Santer BD (1990) Statistical comparison of spatial fields in model validation, perturbation, and predictability experiments. *J Geophys Res* 95: 851-865
- Wigley TML, Jones PD, Kelly PM, Hulme M (1992) Recent global temperature changes: ozone and aerosol influences. *Proceedings 16th Annual Climate Diagnostics Workshop*, pp 194-202
- Wigley TML, Jones PD, Kelly PM, Raper SCB (1989) Statistical significance of global warming. *Proceedings 13th Annual Climate Diagnostics Workshop*, pp A1-A8
- Woodward WA, Gray HL (1993) Global warming and the problem of testing for trend in time series data. *J Climate* 6: 953-962

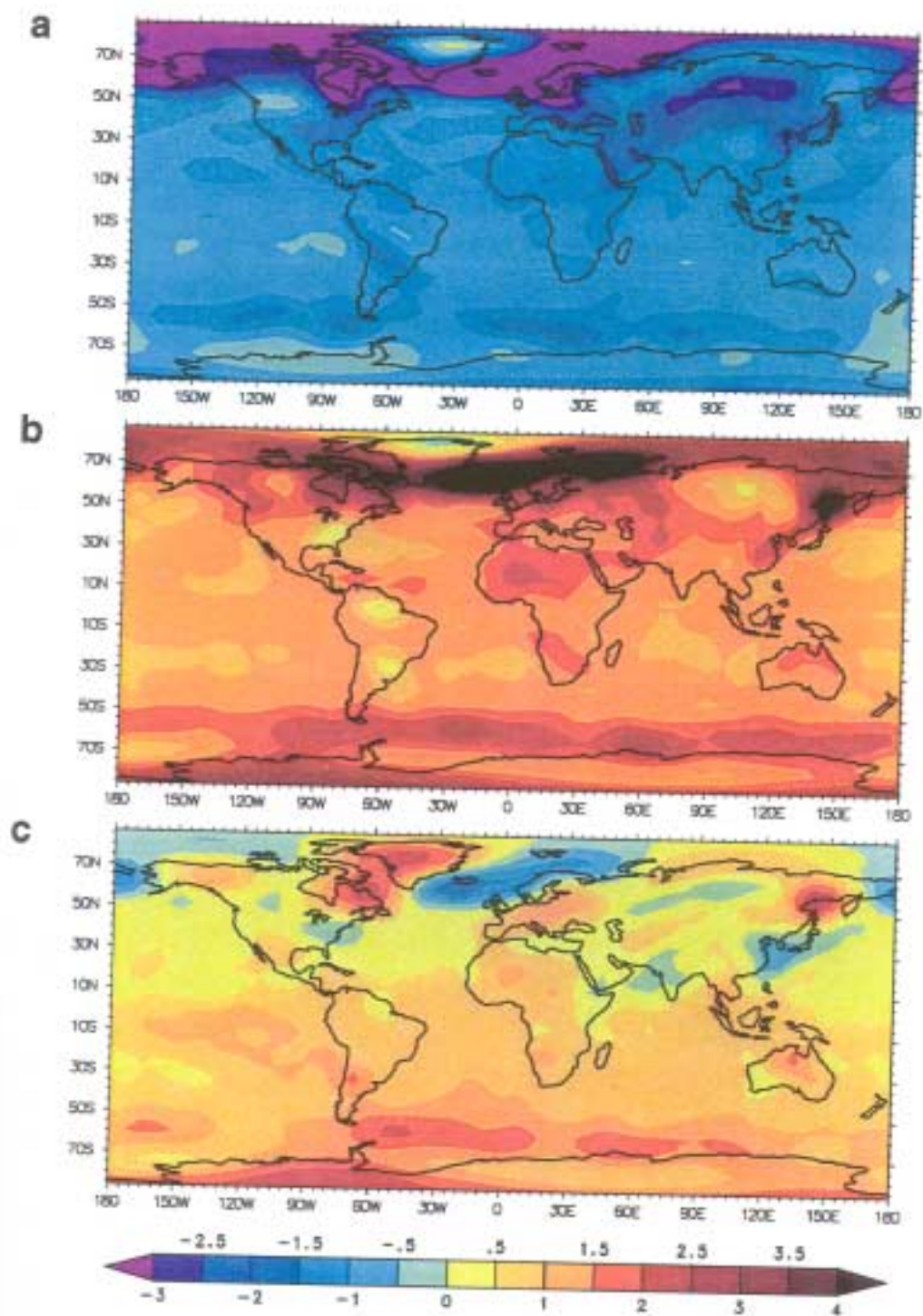


Figure 1: Wintertime (DJF) near-surface temperature changes ($^{\circ}\text{C}$) in the three Taylor and Penner (1994) response experiments, with forcing by anthropogenic sulfate aerosols (1a), CO_2 (1b), and a combination of anthropogenic sulfate aerosols and CO_2 (1c). The perturbations approximately correspond to present-day sulfur emissions and atmospheric CO_2 levels. All changes were computed using 20-year averages and are expressed relative to the 20-year DJF mean of a control run with a nominal pre-industrial CO_2 concentration and no anthropogenic sulfate aerosols.

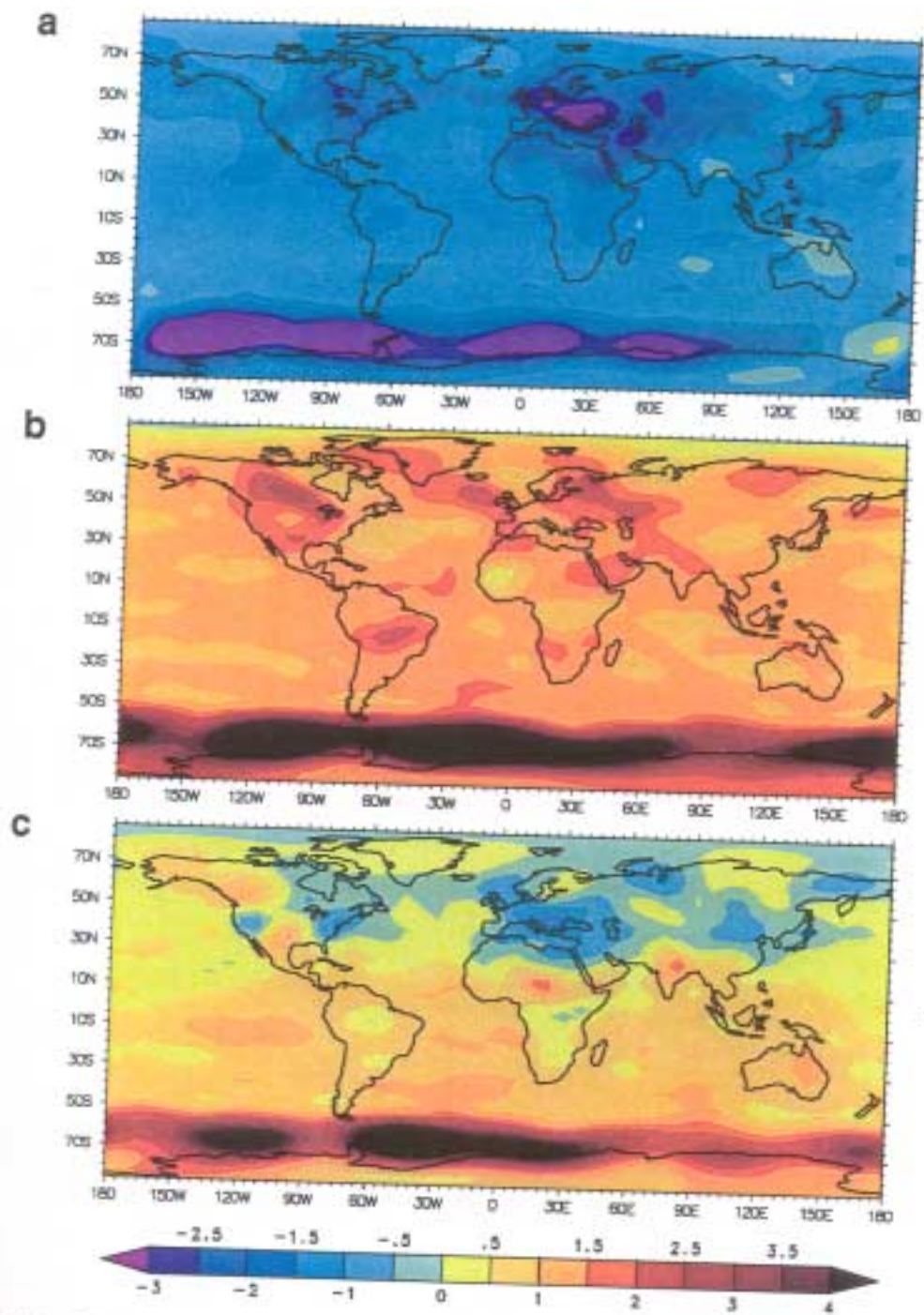


Figure 2: As for Figure 1, but for summertime (JJA) near-surface temperature changes (°C) in the three TP response experiments.

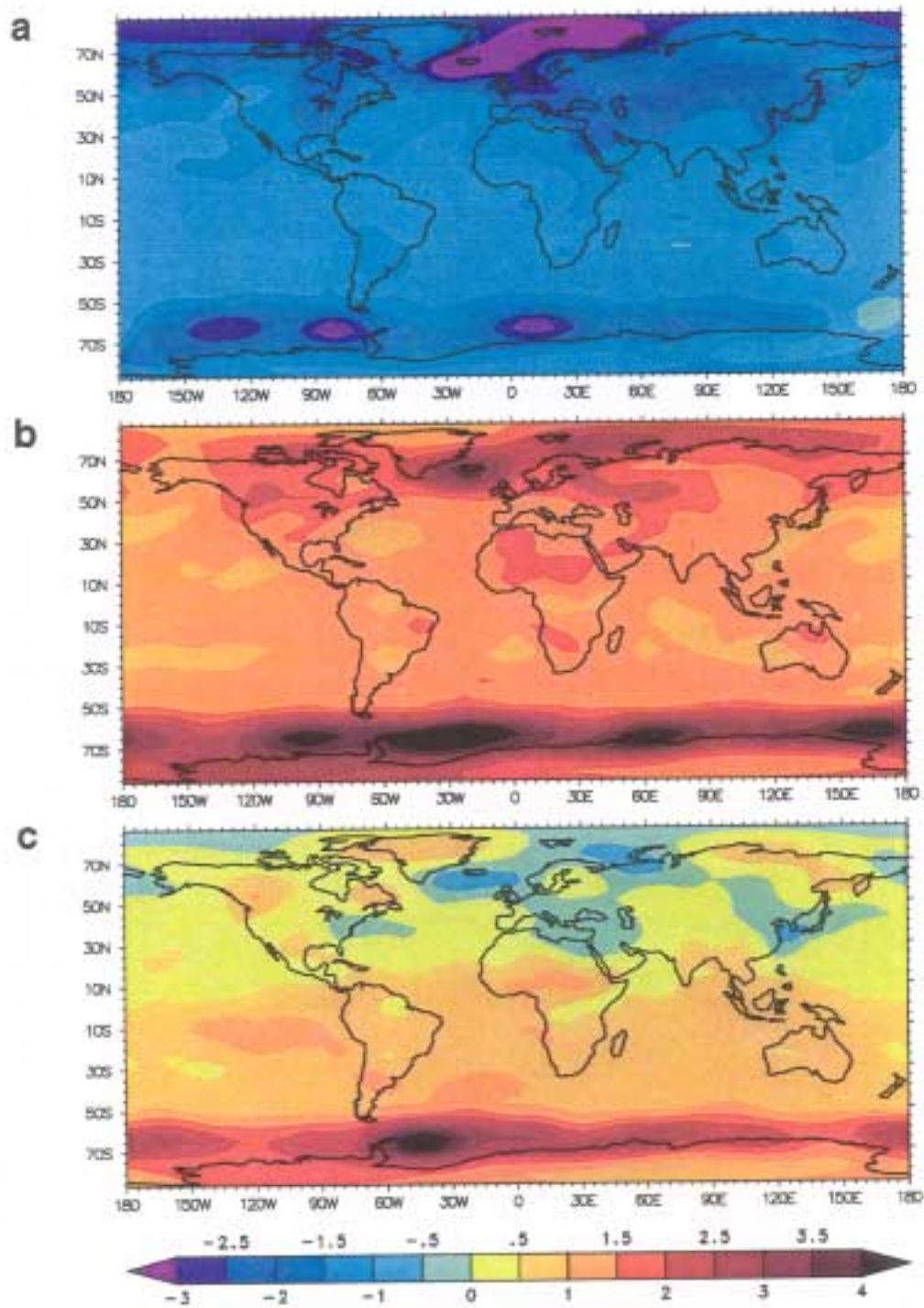


Figure 3: As for Figure 1, but for annually-averaged near-surface temperature changes (°C) in the three TP response experiments.

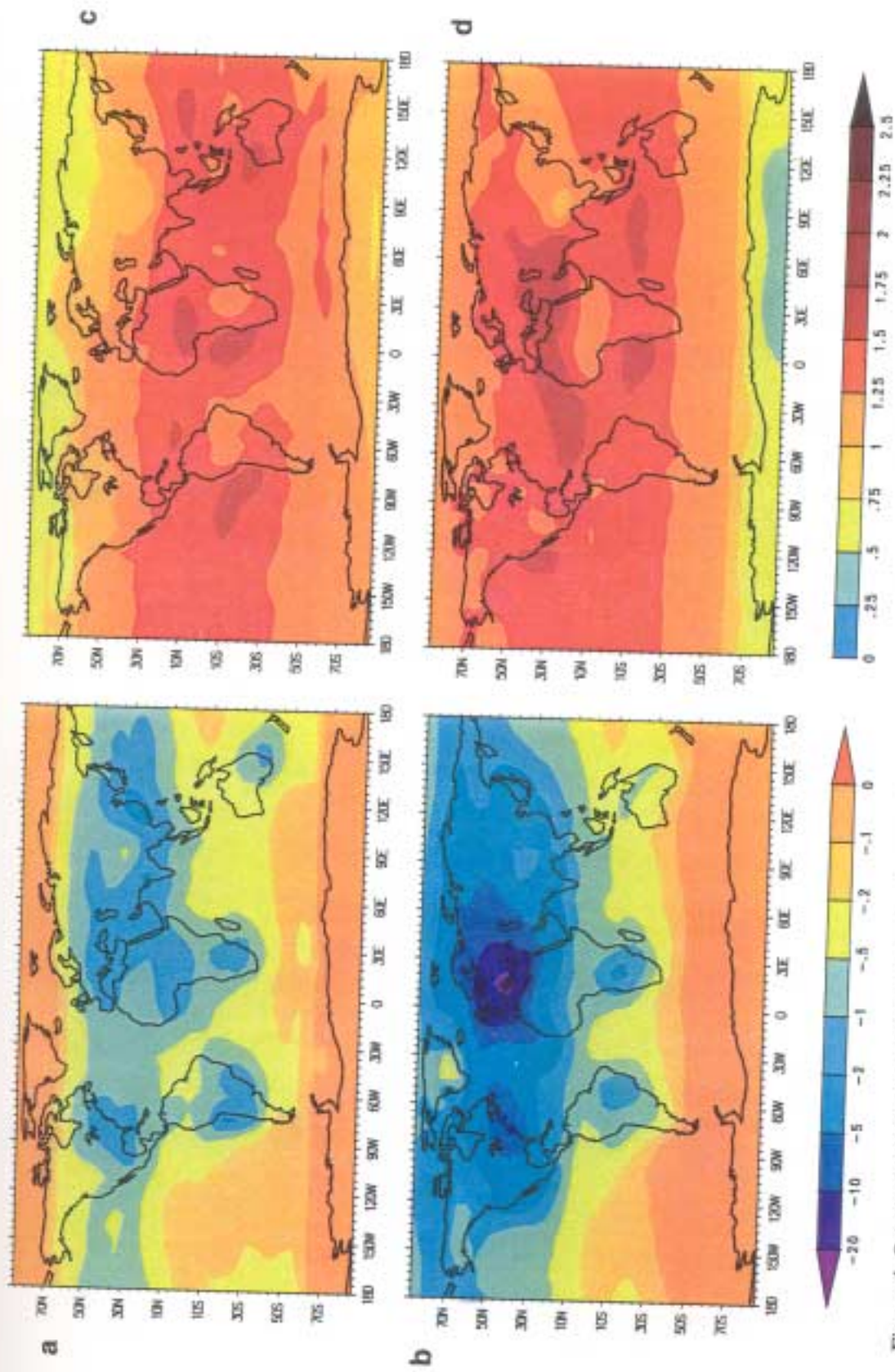


Figure 4: Seasonally-averaged radiative forcing due to sulfate aerosols and CO_2 in the TP sulfate-only and CO_2 -only experiments. The difference in the top-of-the-atmosphere clear-sky outgoing SW radiation (in Wm^{-2}) between simulations with and without anthropogenic sulfate aerosols is shown for winter (DJF; a) and summer (JJA; b). Changes in the net long-wave flux at the tropopause due to an increase in atmospheric CO_2 from 275 ppmv to 345 ppmv are also given for winter (c) and summer (d).

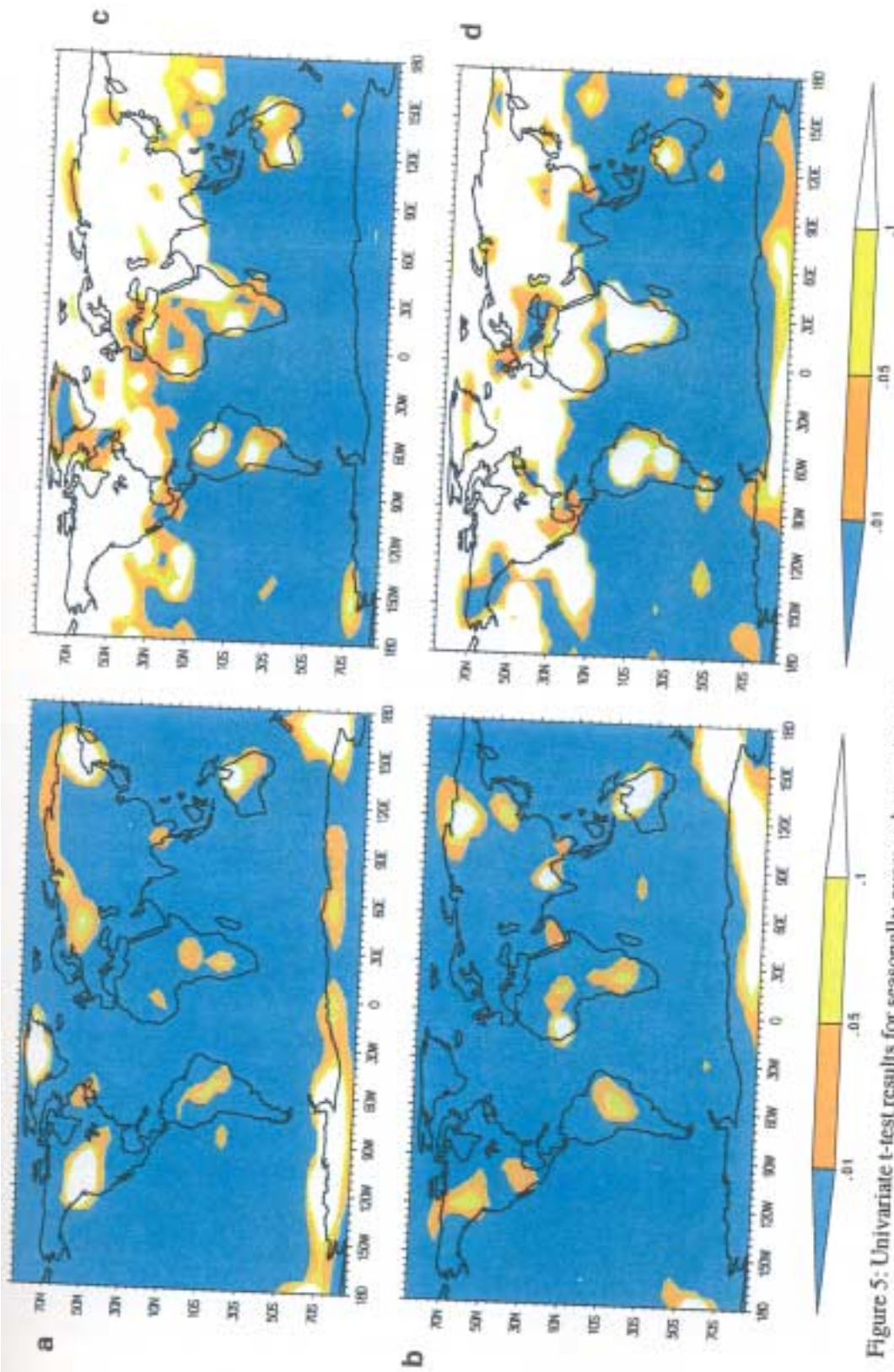


Figure 5: Univariate t-test results for seasonally-averaged near-surface temperature changes ($^{\circ}\text{C}$) in the TP response experiments. Results are for DJF and JJA changes in the sulfate-only experiment (panels a and b, respectively) and in the experiment with combined CO_2 /sulfate aerosol forcing (panels c and d). Areas shaded denote regions with (experiment minus CTL) differences in grid-point time averages which are significant at the 1%, 5% and 10% levels (two-tailed tests). Tests were conducted with 20-year samples of data from the control run and each response experiment. Near-surface temperature changes in the CO_2 -only experiment (not shown) are highly significant at virtually all grid-points.

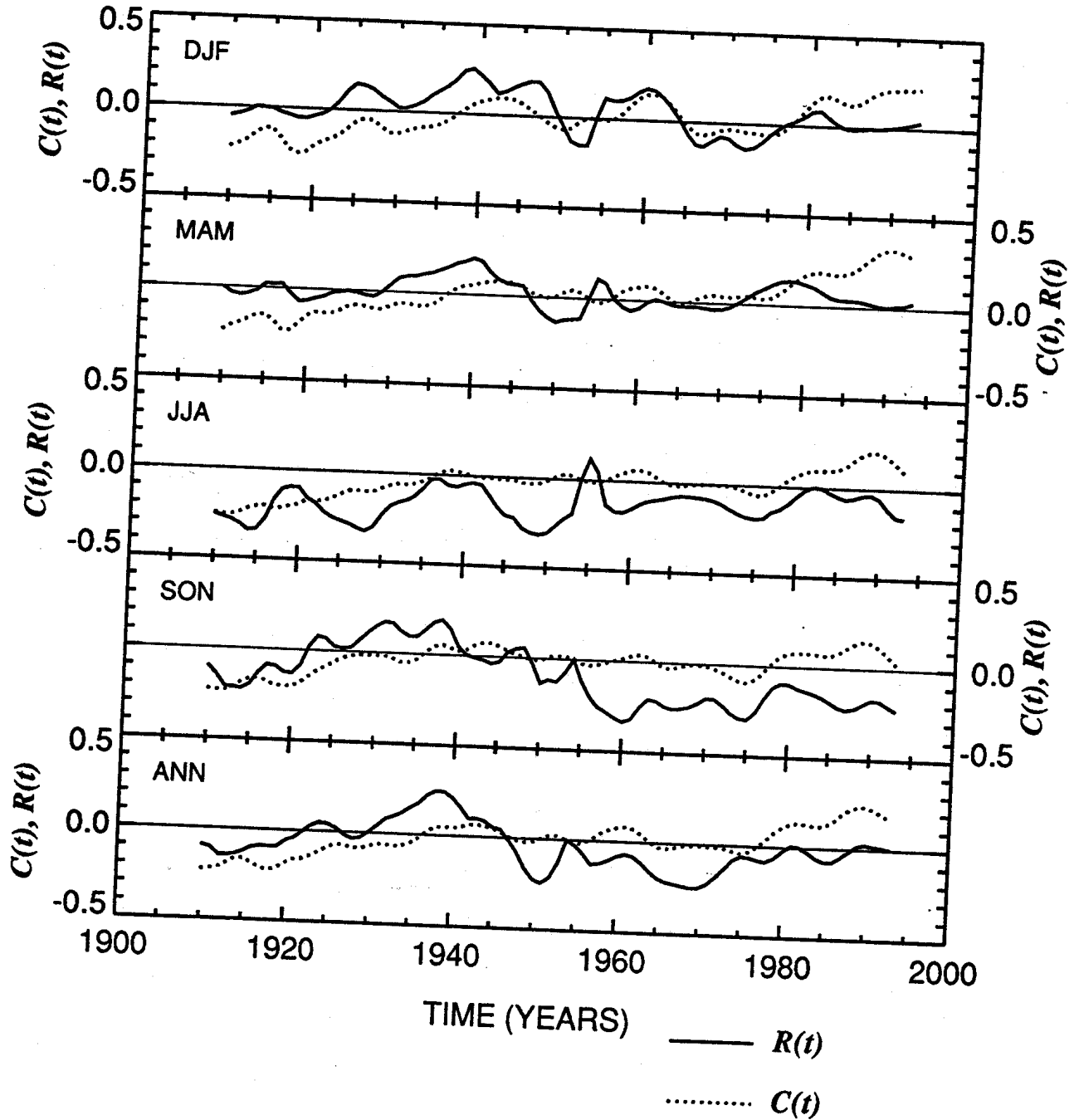


Figure 6: Centered [$R(t)$] and uncentered [$C(t)$] pattern correlations between model and observed near-surface temperature changes. Model changes in seasonally- and annually-averaged temperature are taken from the the TP CO₂-only experiment. For each season, one time-independent spatial pattern characterizes the response to forcing by present-day atmospheric CO₂ concentrations. Observed changes were filtered to remove high-frequency noise, and were expressed as a time series of 84 anomaly patterns (from 1910-1993) relative to a filtered reference period extending from 1948-1960. All pattern correlations were calculated after excluding grid-points with missing data.

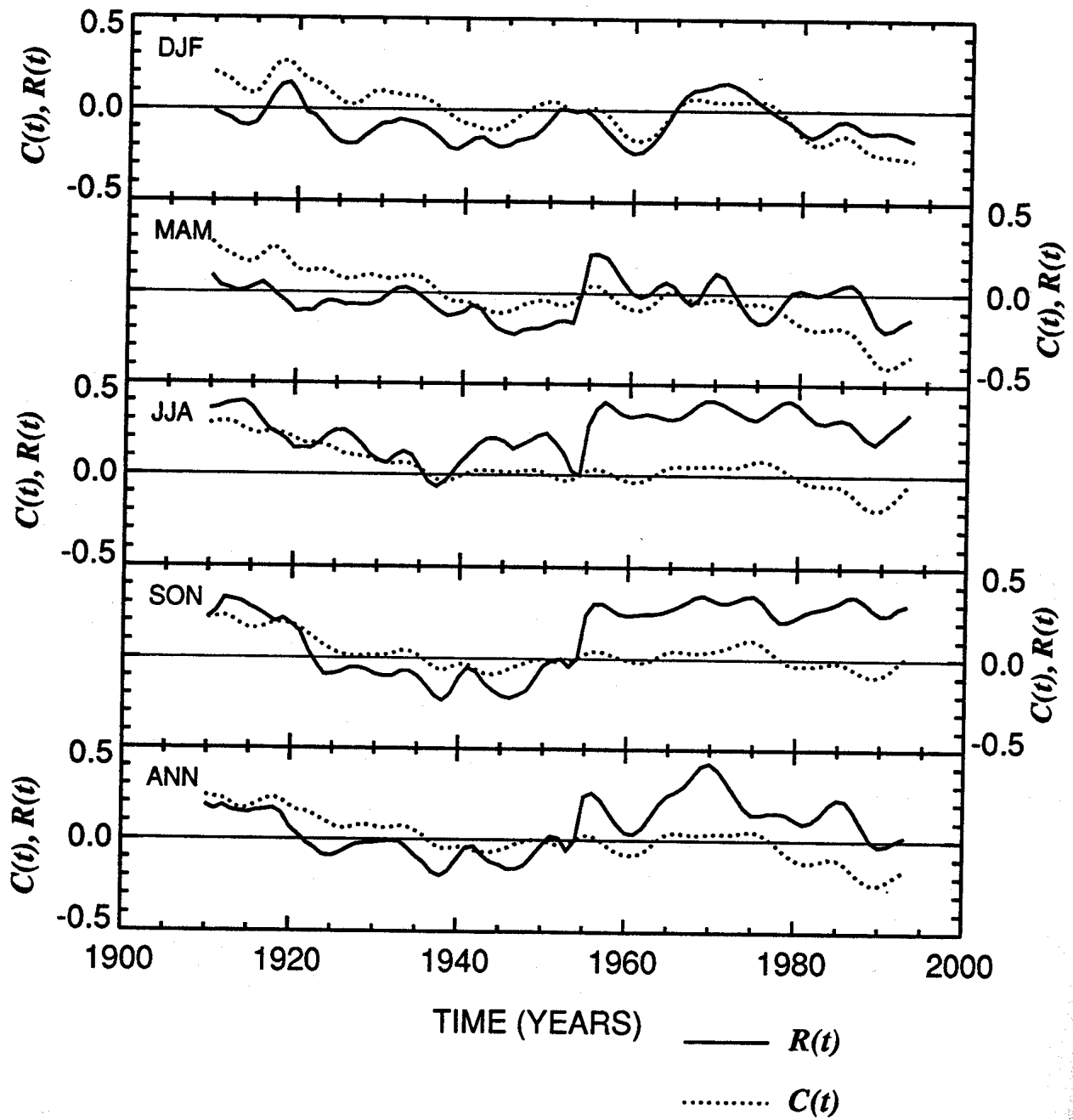


Figure 7: As for Figure 6, but for near-surface temperature changes from the TP sulfate-only experiment.

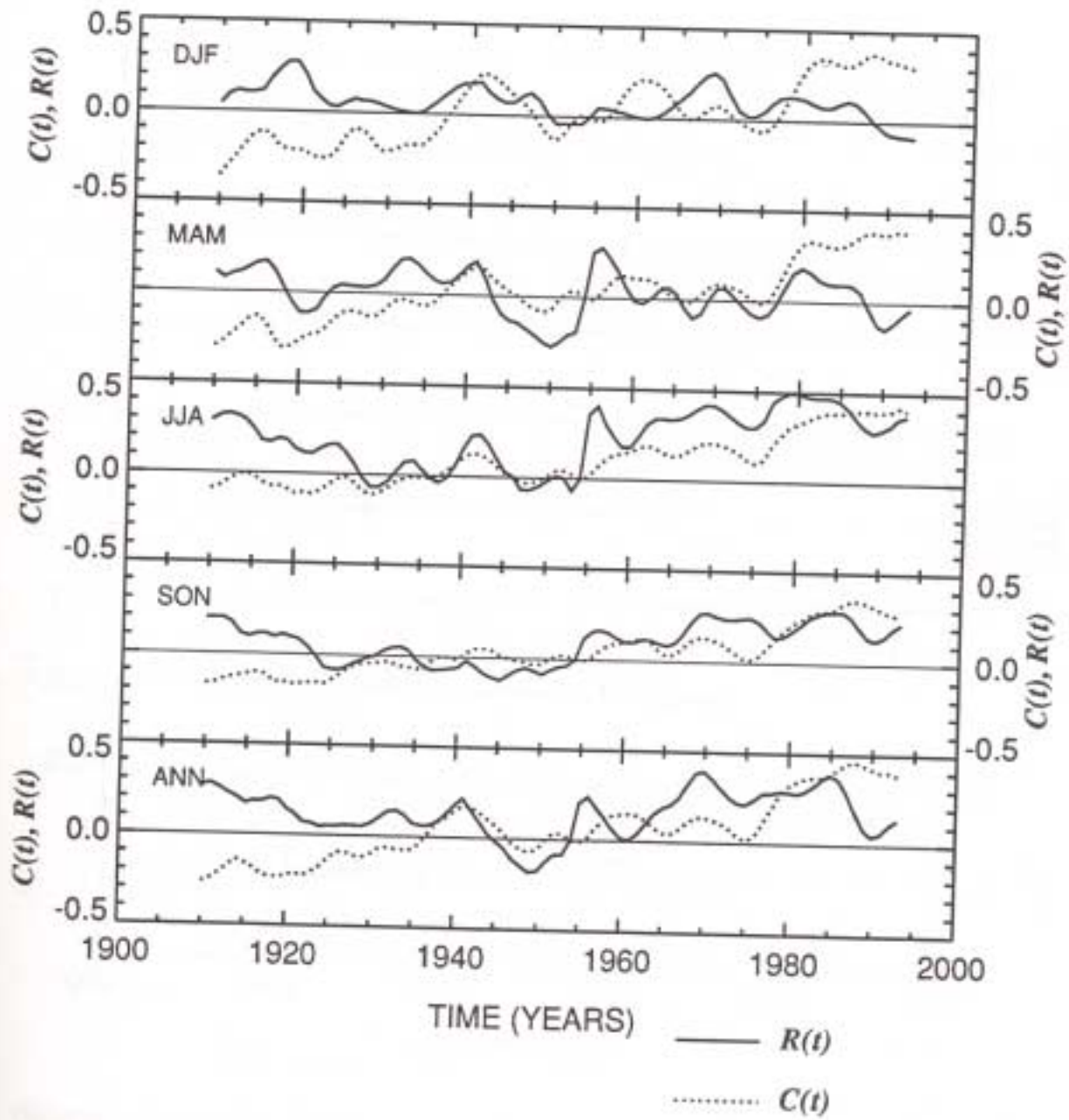


Figure 8: As for Figure 6, but for near-surface temperature changes from the TP experiment with combined CO_2 and sulfate aerosol forcing.

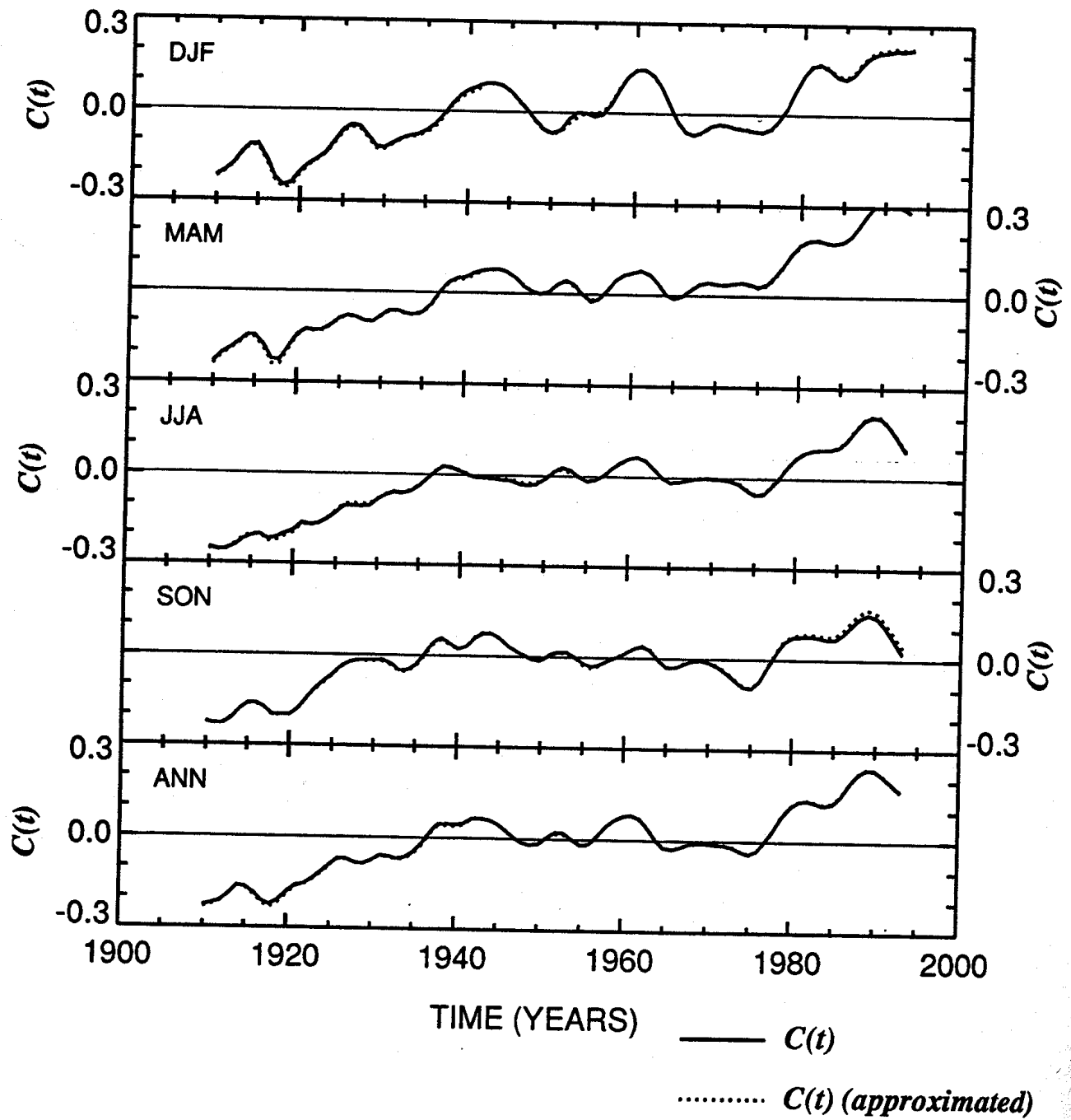


Figure 9: Actual and approximated $C(t)$ time series for near-surface temperature changes from the TP CO_2 -only experiment. Approximated values of $C(t)$ were computed using equation (8) and are very nearly obscured by the actual values.

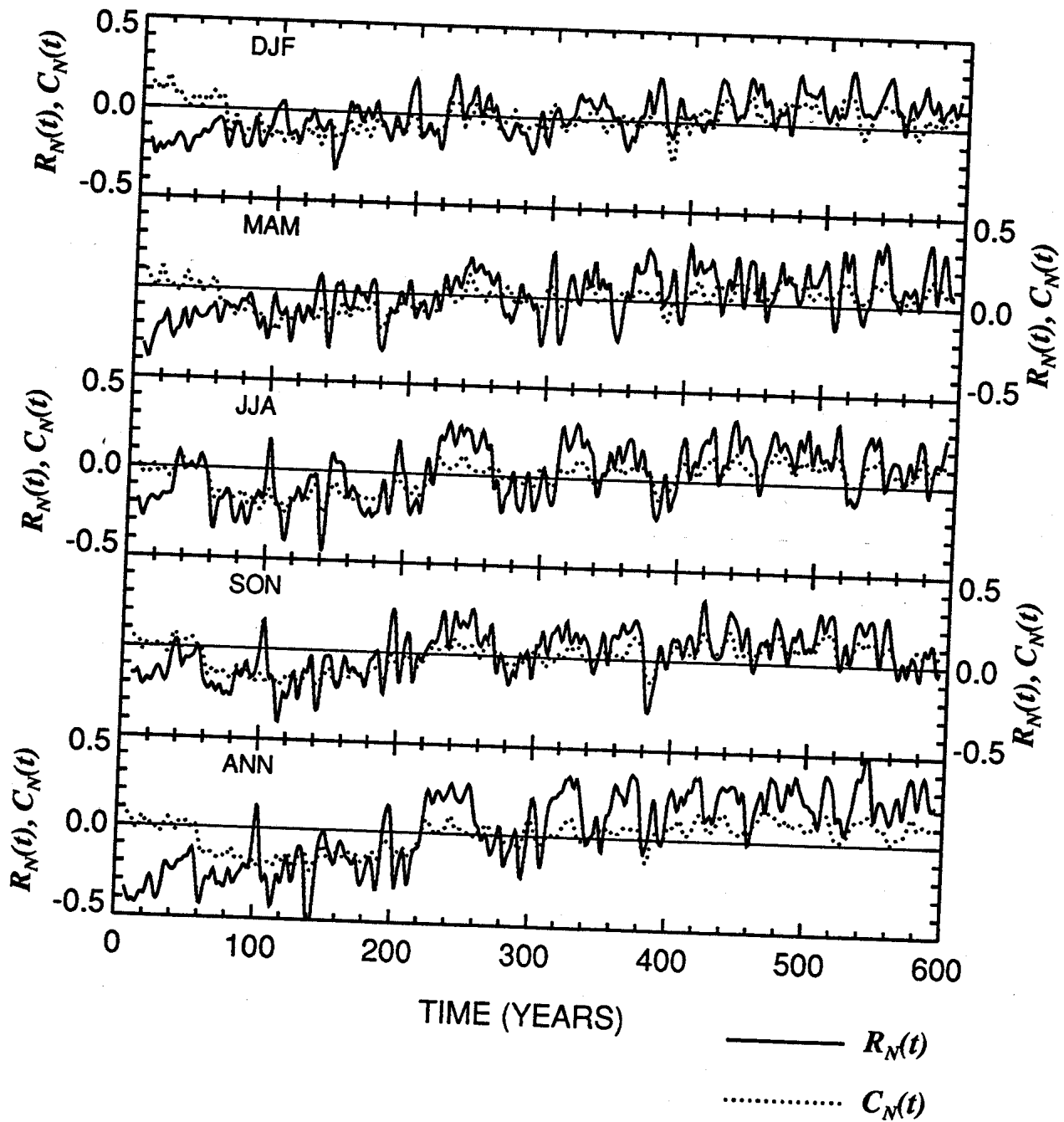


Figure 10: Behavior of pattern correlation statistics in the absence of external forcing. $R_N(t)$ and $C_N(t)$ time series were computed using (filtered) time-dependent patterns of near-surface temperature change from a 600-year control integration ("HAMCTL") with a fully-coupled A/OGCM and time-independent seasonal- and annual temperature-change patterns from the TP experiment with combined CO_2 / sulfate aerosol forcing.

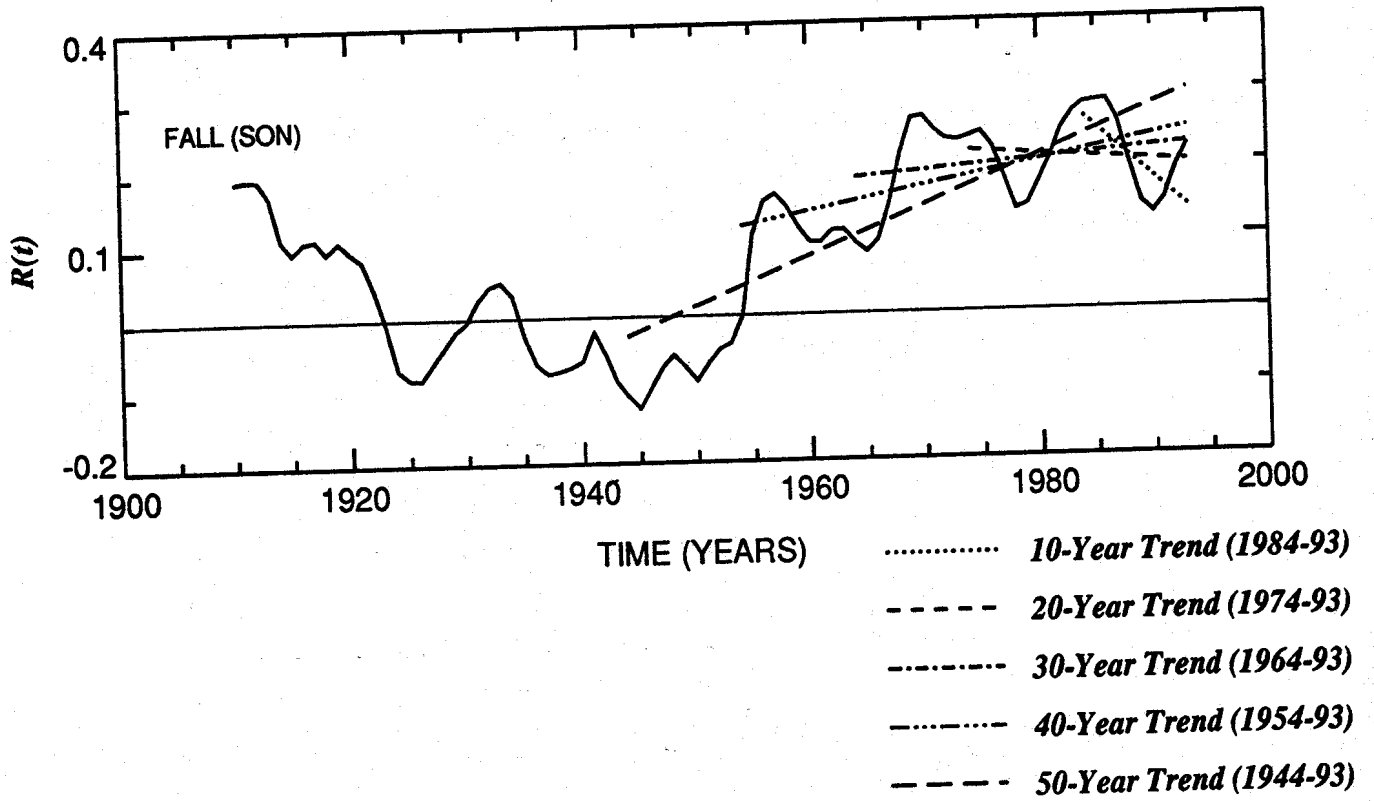


Figure 11: Linear trends for the final 10-50 years of the $R(t)$ time series for the SON near-surface temperature signal from the TP experiment with combined CO_2 / sulfate aerosol forcing. These are the signal trends $\beta_S(i)$, whose significance we wish to determine.

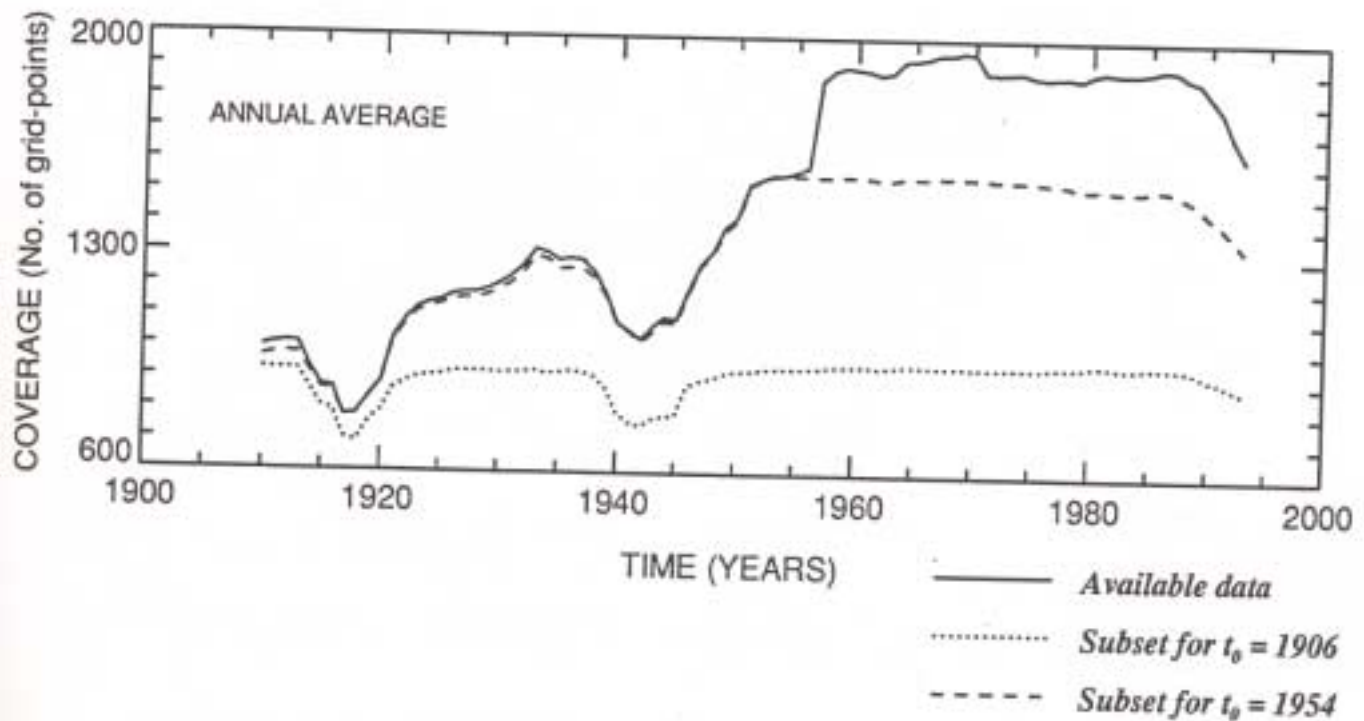


Figure A1: Observed changes in coverage of gridded, annually-averaged near-surface temperature data as a function of time. The curve labelled "Available data" shows the change in coverage for the filtered Jones et al. (1991) data if the data are used "as is" (i.e., as anomalies relative to 1950-79). The curves labelled " $t_0 = 1906$ " and " $t_0 = 1954$ " show the coverage changes which are obtained by requiring that the data coverage in $\bar{D}_2(x, t)$ is a subset of the coverage in $\bar{D}_1(x)$ (see Appendix). For the earlier reference period, the coverage fluctuates around ca. 925 grid-points from ca. 1910-1988, with minima of ca. 700 and 750 grid-points in the war years (due to a reduction in oceanic data availability). The coverage is less stable with time for $t_0 = 1954$, since hundreds of grid-points that had data in the reference period 1948-60 did not have data in the first few decades of the century. Note, however, that the average coverage from ca. 1950-90 (roughly 1,550 grid-points) is much higher than for $t_0 = 1906$. In all three curves, recent reductions in coverage reflect the degradation of the station observing network over land.

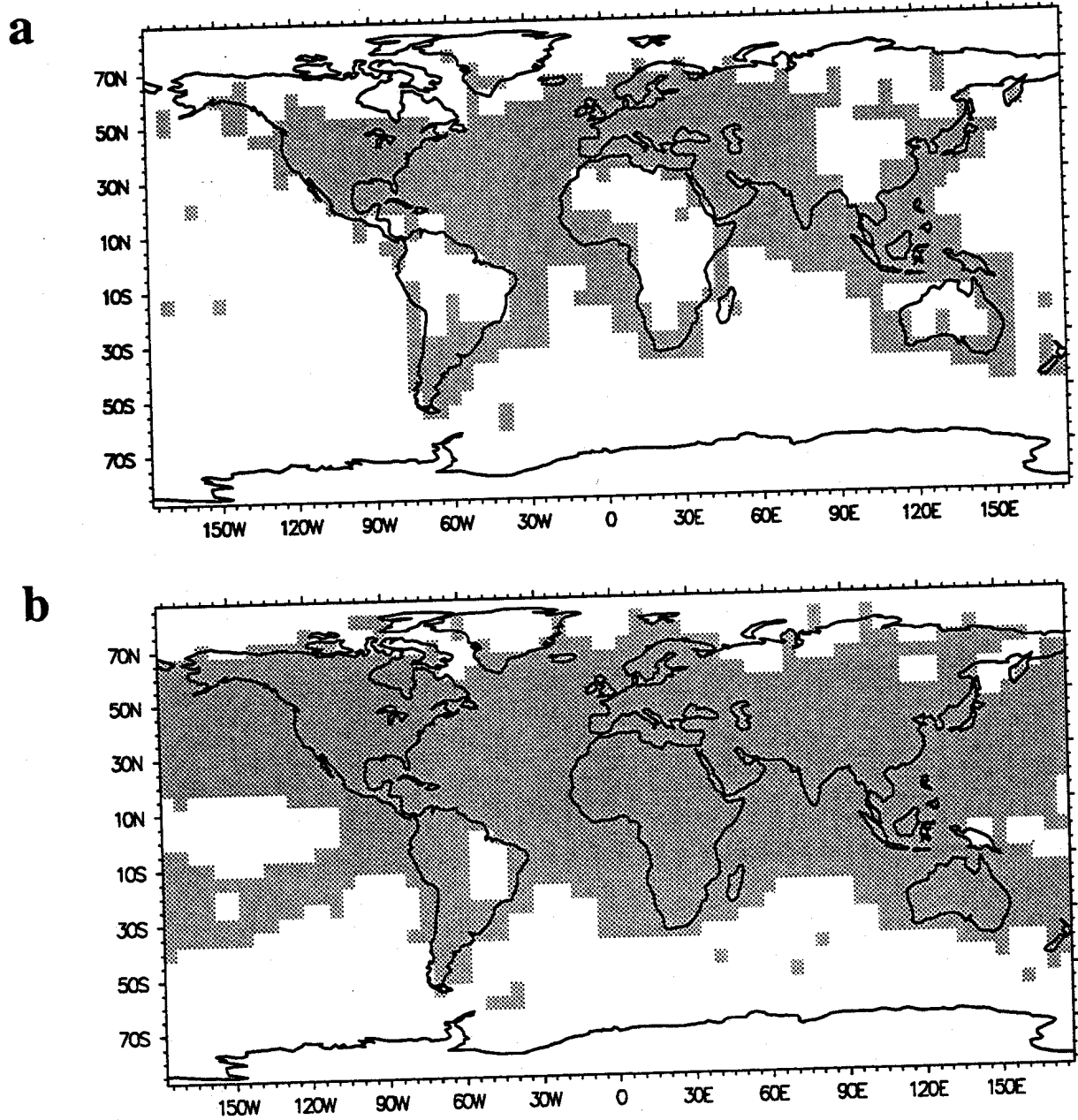


Figure A2: Spatial coverage of annual mean near-surface temperature data for two different reference periods: $t_0 = 1906$ (panel a) and $t_0 = 1954$ (panel b).

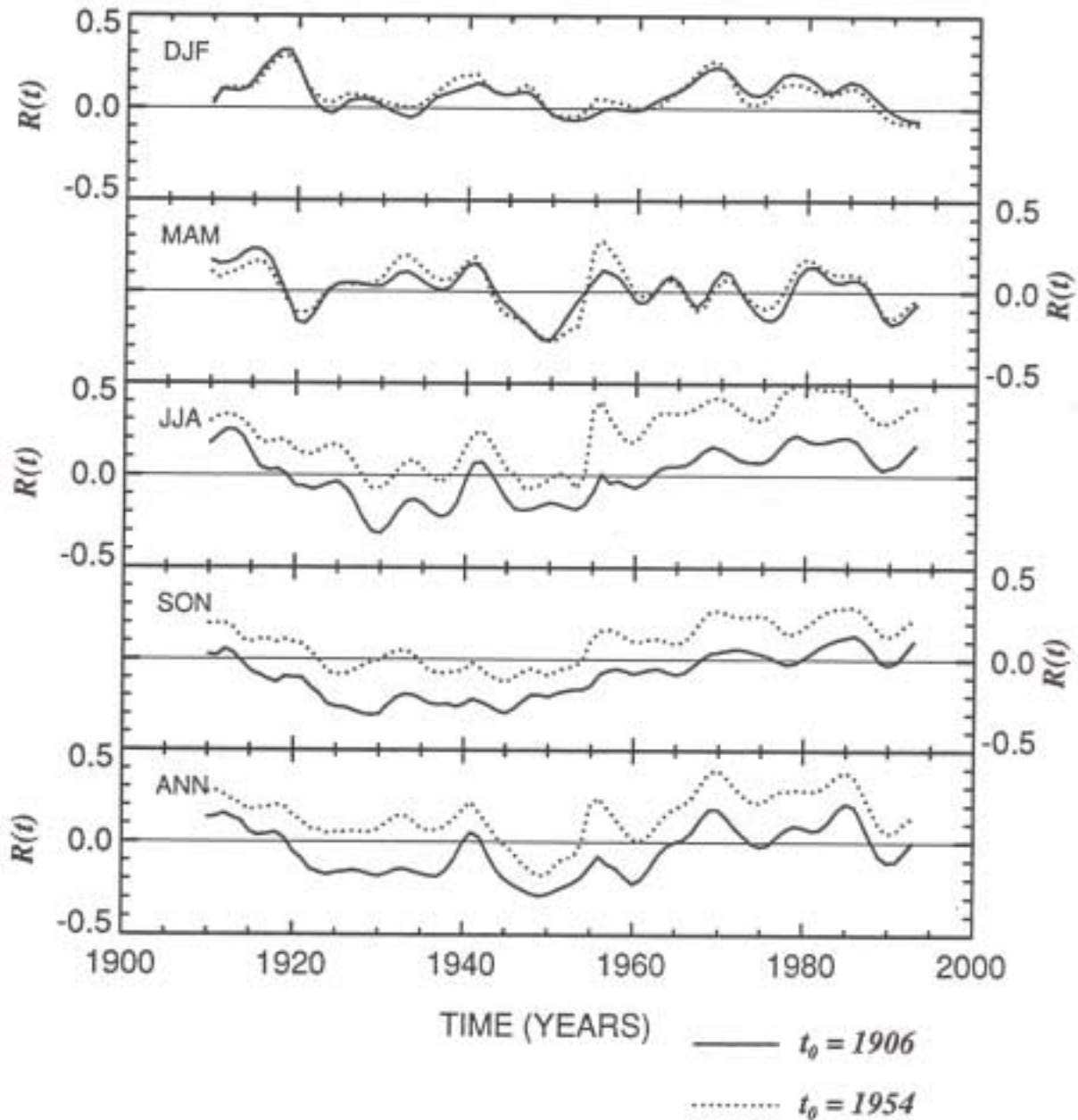


Figure A3: Centered pattern correlation, $R(t)$, between observed seasonally- and annually-averaged near-surface temperature changes and changes in the TP response experiment with combined sulfate aerosol / CO_2 forcing. Results are for two different choices of reference year: $t_0 = 1906$ and $t_0 = 1954$. Changes in $R(t)$ with time are relatively insensitive to the choice of reference year. Note that the curves are forced to pass through zero at t_0 .

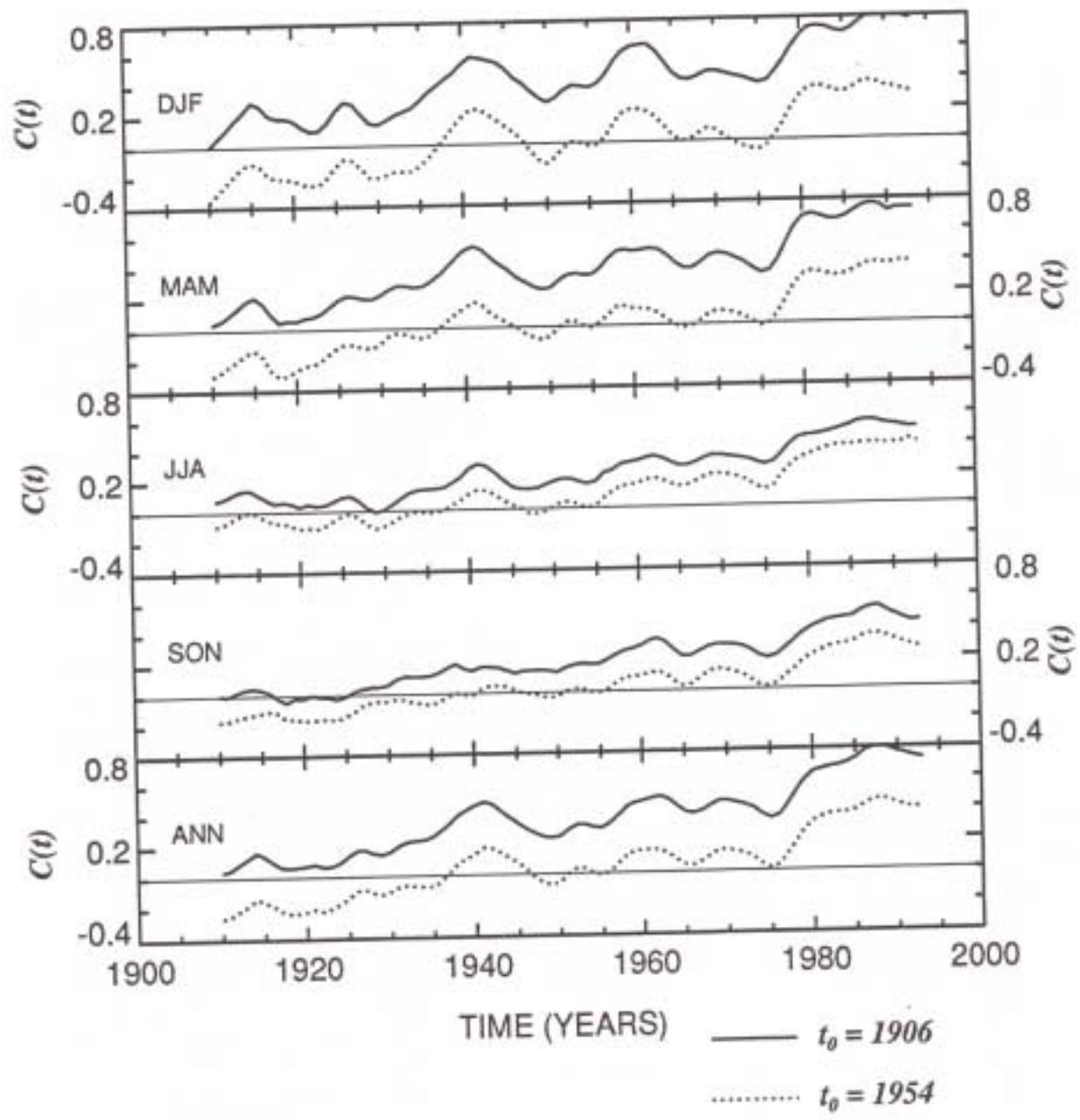


Figure A4: As for Figure A3, but for the uncentered pattern correlation $C(t)$, without subtraction of model and observed spatial means.

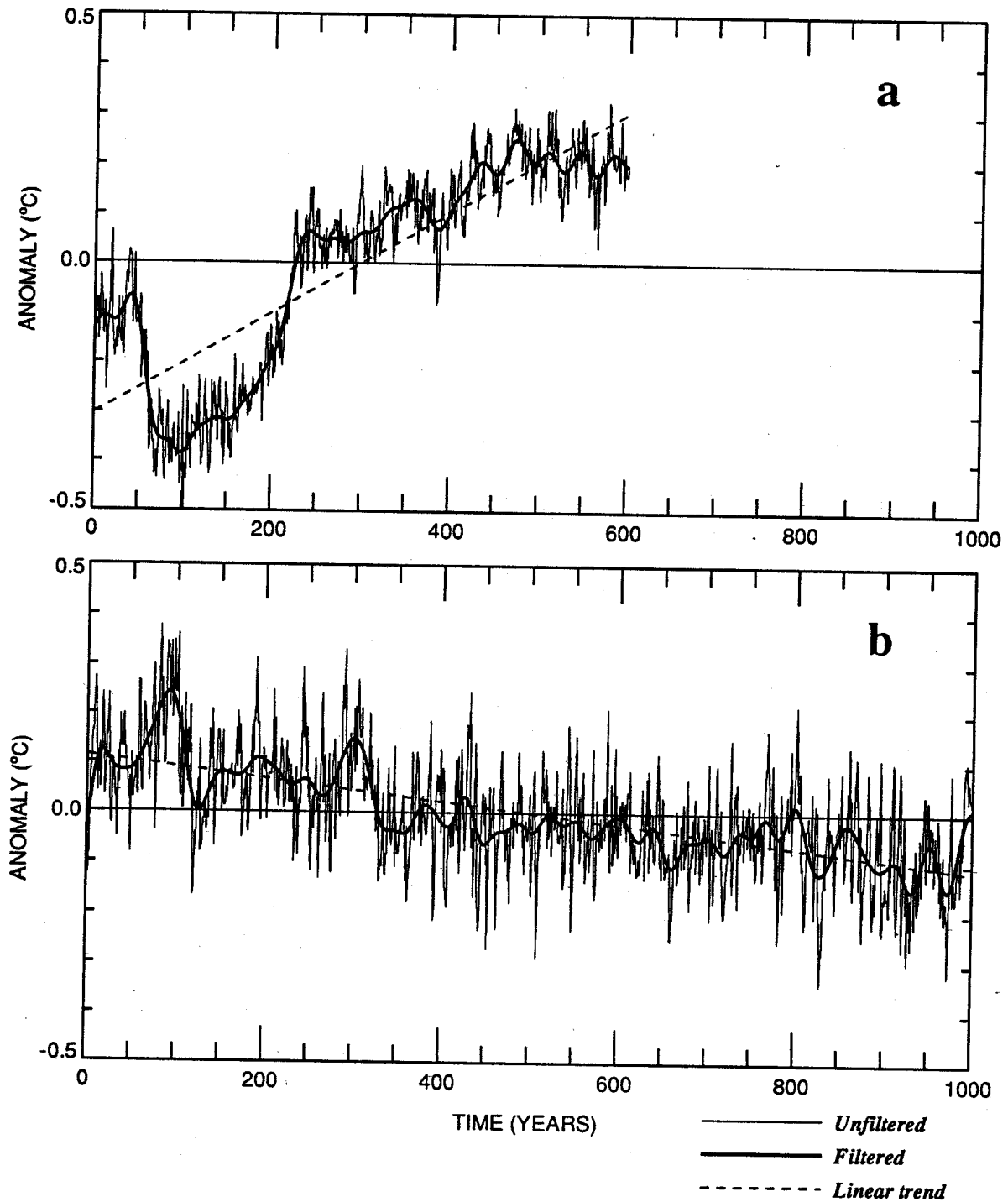


Figure B1: Time series of global mean, annually-averaged near-surface temperature changes in the 600-year HAMCTL (panel a) and 1,000-year GFDLCTL integrations (panel b). Both control runs were performed with fully-coupled A/OGCMs. Anomalies were defined relative to the mean state over the entire integration. The unfiltered and low-pass filtered time series are shown, together with the least-squares linear trends. Both time series are non-stationary: the linear trend per century is roughly four times larger in HAMCTL than in GFDLCTL (0.102°C versus 0.023°C , respectively).

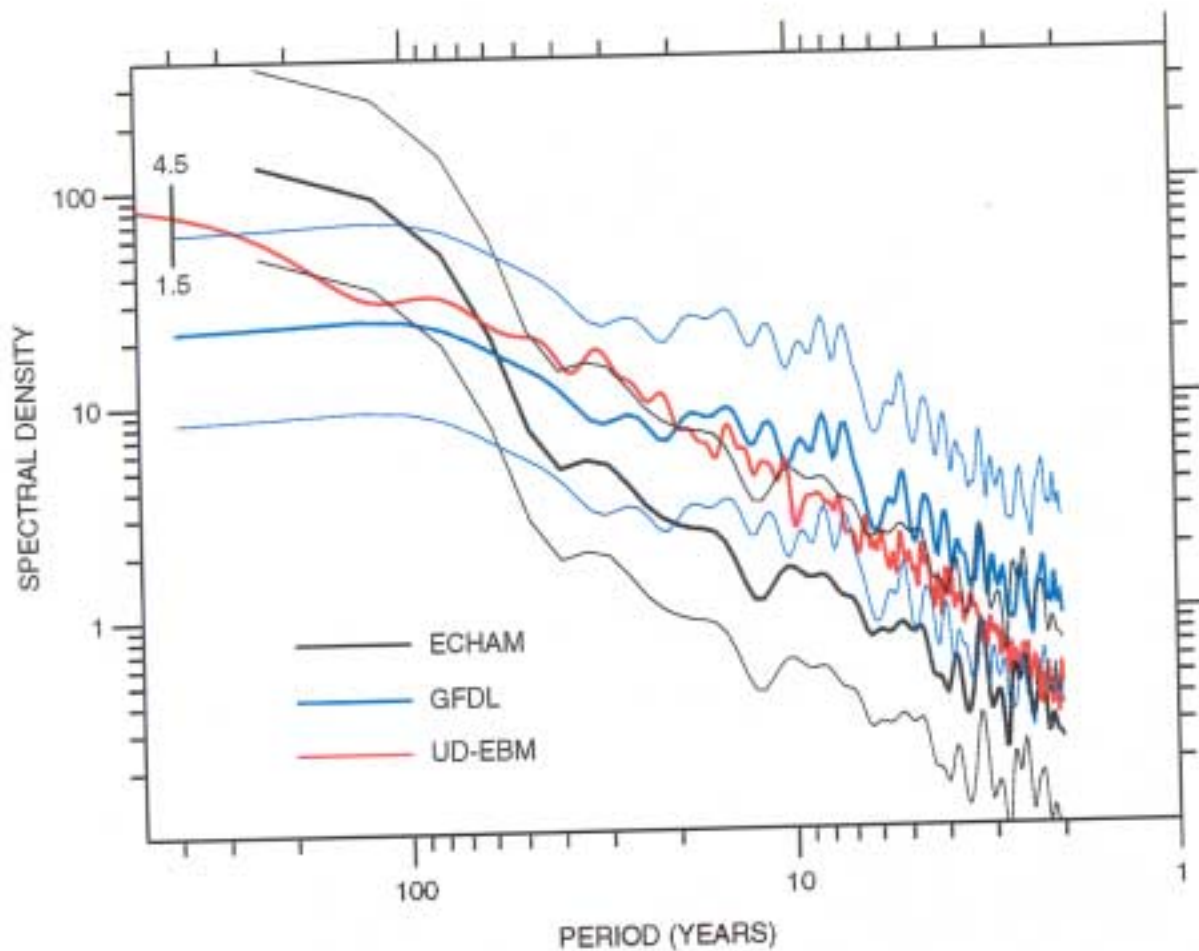


Figure B2: Power spectrum of global mean, annually-averaged near-surface temperature changes in the 600-year HAMCTL and 1,000-year GFDLCTL integrations, and in a 5,000-year integration performed with an upwelling-diffusive EBM (Wigley and Raper, 1991b). Spectra were computed by taking the Fourier transform of the autocorrelation function, with the maximum number of lags equal to 1/4 the length of the time series (see Jenkins and Watts, 1968). Spectra were smoothed using a Tukey window, with window widths of 60, 100, and 200 lags (for HAMCTL, GFDLCTL, and the EBM, respectively). The number of frequencies computed is twice the window width (for HAMCTL and GFDLCTL) and four times the window width for the EBM. The thin lines are the 95% confidence intervals for the coupled model spectral estimates. HAMCTL has higher power than GFDLCTL for periods longer than roughly 60 years; the reverse is true for periods shorter than 60 years. The EBM was run with random forcing and a climate sensitivity of 3.0°C for a doubling of CO_2 . The bar on the upper left corner of the figure gives the EBM spectral densities for climate sensitivities of 1.5°C and 4.5°C at a period of 400 years. The coupled model spectra were computed after removal of overall linear trends in the data.

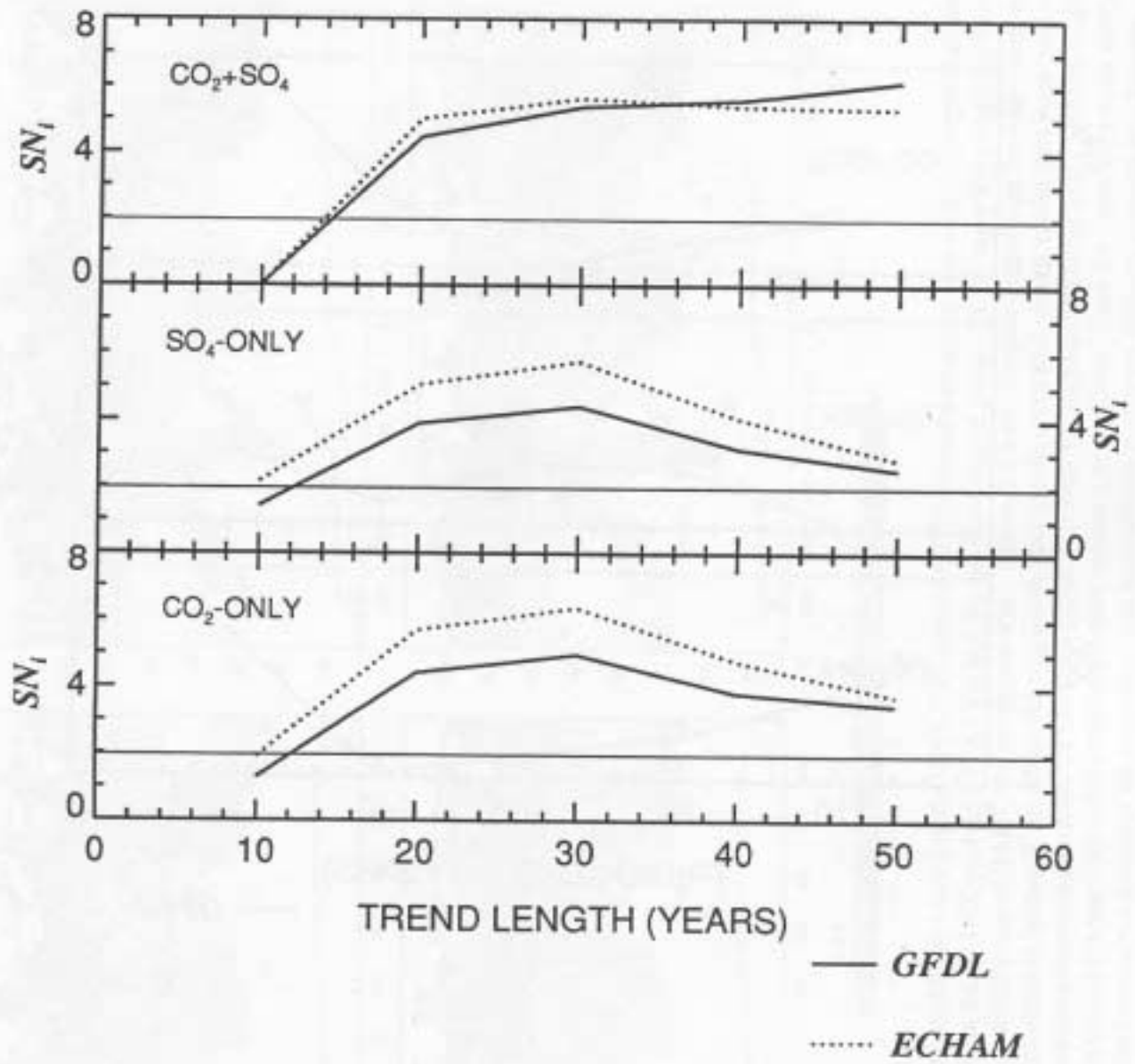


Figure B3: Signal-to-noise ratios, SN_i , for observed trends in the uncentered pattern correlation statistic $C(t)$. The signals, $\beta_S(i)$, are the observed trends in $C(t)$ for the last 10- to 50 years. The noise, $\epsilon_N(i)$, is the standard error of the sampling distribution of $\beta_N(i)$ (see Section 6.1 and Appendix B for further details). Standard errors were computed using data from both GFDLCTL and HAMCTL. Results are for annually-averaged near-surface temperature data. For temperature-change signals from the Taylor and Penner sulfate-only and CO₂-only experiments, SN_i is consistently lower for the GFDLCTL noise estimates. The GFDL versus ECHAM differences in SN_i can be interpreted in terms of the coupled-model differences in the spectral density of global-mean, annually-averaged temperature on timescales of 10- to 50 years (see Figure B2). The thin solid line gives the 5% significance threshold for an assumed Gaussian distribution of trends.

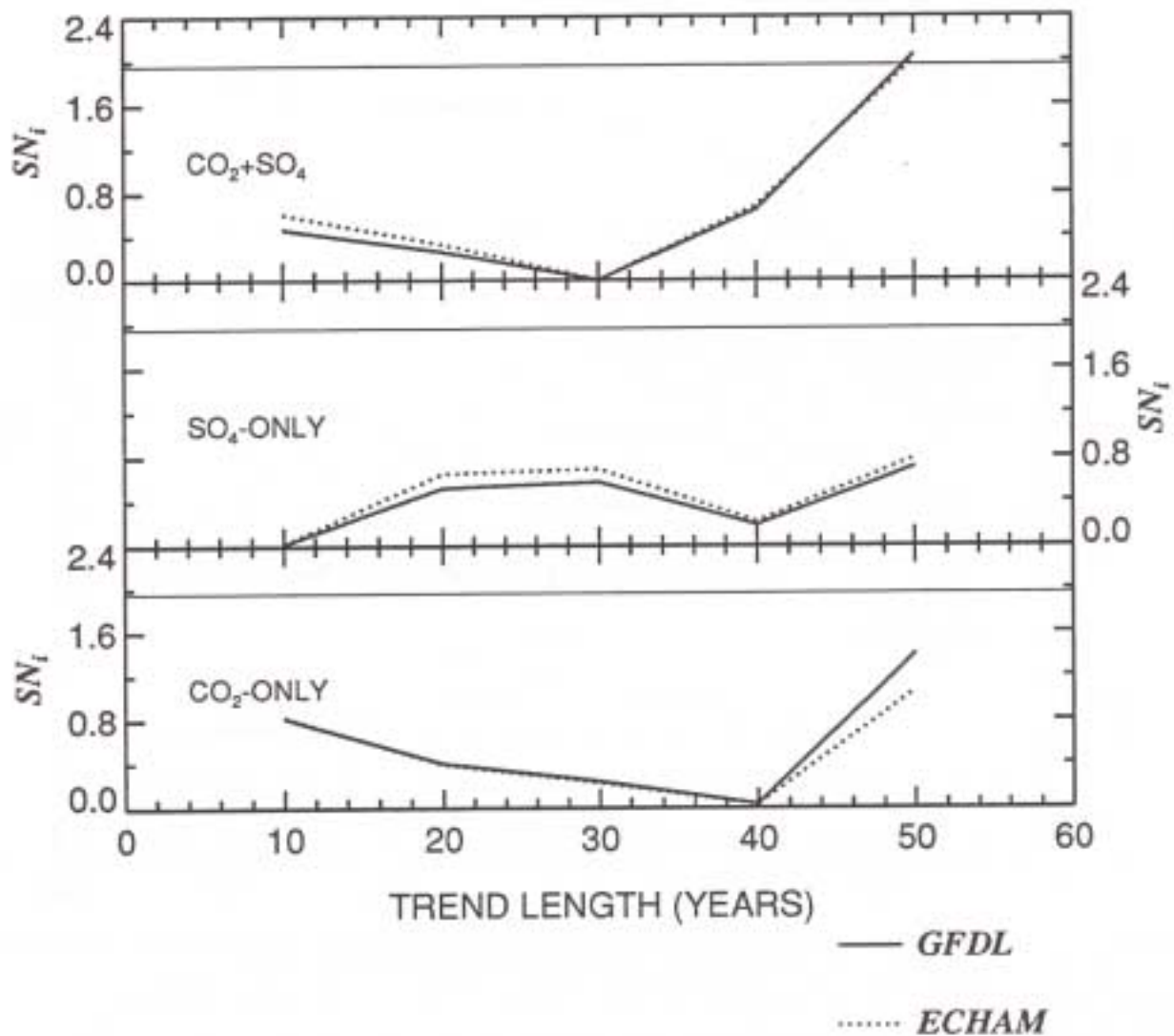


Figure B4: As for Figure B3, but for observed trends in the centered pattern correlation statistic $R(t)$ and for seasonally-averaged data (JJA). SN_i ratios are very similar for the two noise estimates. This result implies some similarity in the HAMCTL and GFDLCTL patterns of near-surface temperature variability over the observed data window on 10- to 50-year timescales.

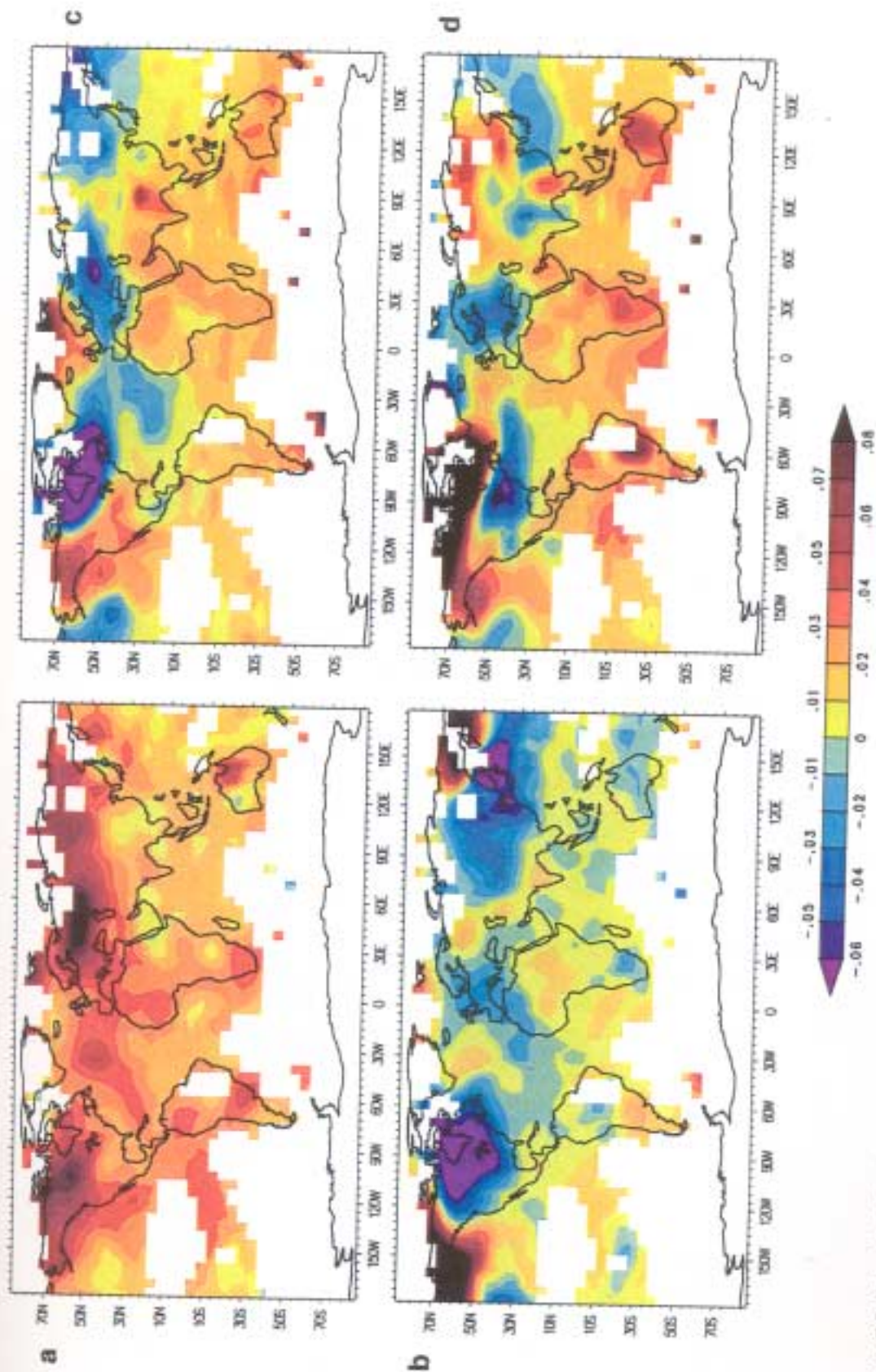


Figure B5: Empirical Orthogonal Functions (EOFs) 1 and 2 of the GFDLCTL and HAMCTL annually-averaged near-surface temperature anomaly data. EOFs were computed using data from the first 400 years of GFDLCTL and the last 400 years of HAMCTL (in order to exclude the large negative temperature excursion during the first 200 years of HAMCTL; see Figure B1a). Anomalies were computed relative to the 400-year means of each control run, and were filtered as described in Section 4. To facilitate comparison with the observed data mask for 1954 (see Figure A2) results, model data were interpolated to the observed grid, and grid-points outside the observed data mask for 1954 (see Figure A2) were not used for computing EOFs. GFDLCTL (HAMCTL) EOFs 1 and 2 explain 13.2% and 8.9% (11.0% and 9.9%) of the total space-time variance, respectively. There are some large-scale similarities in the EOF patterns: both coupled models have large loadings of the same sign over high-latitude land areas of North America and Eurasia.

SEASON	FORCING	RESPONSE		
		SO ₄	CO ₂	SO ₄ + CO ₂
DJF	SO ₄	0.02	0.55	0.27
	CO ₂	0.20	-0.51	-0.03
	SO ₄ + CO ₂	0.30	0.15	0.37
JJA	SO ₄	0.36	0.27	0.38
	CO ₂	0.08	-0.52	0.04
	SO ₄ + CO ₂	0.49	-0.52	0.46
ANN	SO ₄	0.36	0.51	0.43
	CO ₂	0.09	-0.56	0.01
	SO ₄ + CO ₂	0.52	-0.22	0.49

Table 1: Global spatial pattern correlations between forcing and response fields. The forcing fields are either the seasonally- or annually-averaged radiative forcing due to the individual and combined effects of sulfate aerosols and CO₂ (see Figure 4 for details). The response fields are the time-averaged near-surface temperature changes in Figures 1-3. All temperature changes were computed with 20-year samples relative to the average over years 11-30 of the TP control integration with no anthropogenic sulfate aerosols and pre-industrial atmospheric CO₂ levels. All correlations are for an area-weighted form of R (spatial mean subtracted).

<i>SEASON</i>	<i>EXPT</i>	$\alpha = 0.01$	$\alpha = 0.05$
<i>DJF</i>	<i>SO₄</i>	86.5	93.0
	<i>CO₂</i>	91.8	95.8
	<i>SO₄ + CO₂</i>	56.1	65.3
<i>MAM</i>	<i>SO₄</i>	89.4	93.7
	<i>CO₂</i>	89.6	95.4
	<i>SO₄ + CO₂</i>	55.2	65.7
<i>JJA</i>	<i>SO₄</i>	88.3	93.4
	<i>CO₂</i>	90.0	94.4
	<i>SO₄ + CO₂</i>	55.5	64.5
<i>SON</i>	<i>SO₄</i>	87.2	92.6
	<i>CO₂</i>	95.8	98.3
	<i>SO₄ + CO₂</i>	54.1	65.0
<i>ANN</i>	<i>SO₄</i>	98.8	99.2
	<i>CO₂</i>	100.0	100.0
	<i>SO₄ + CO₂</i>	65.6	71.1

Table 2: Univariate t-test results for differences in mean state between the TP control run with no anthropogenic sulfate aerosols and pre-industrial atmospheric CO₂ and the three TP perturbation experiments with individual and combined CO₂ and sulfate aerosol forcing. All tests are two-tailed, and use 20-year samples of seasonally- and annually-averaged grid-point near-surface temperature data from the control and perturbation experiments. Results indicate the number of local (grid-point) rejections of the null hypothesis of no difference in means (at stipulated significance levels of $\alpha = 0.01, 0.05$), expressed as a percentage of the total number of tests performed.

<i>SEASON</i>	<i>EXPT</i>	<i>CO₂</i>		<i>SO₄ + CO₂</i>	
<i>DJF</i>	<i>SO₄</i>	-0.64	<i>-0.73</i>	0.46	<i>0.48</i>
	<i>CO₂</i>			0.01	<i>-0.18</i>
<i>MAM</i>	<i>SO₄</i>	-0.53	<i>-0.31</i>	-0.26	<i>0.49</i>
	<i>CO₂</i>			0.79	<i>0.29</i>
<i>JJA</i>	<i>SO₄</i>	-0.59	<i>-0.30</i>	-0.26	<i>0.58</i>
	<i>CO₂</i>			0.79	<i>0.32</i>
<i>SON</i>	<i>SO₄</i>	-0.38	<i>-0.40</i>	0.39	<i>0.59</i>
	<i>CO₂</i>			0.52	<i>0.20</i>
<i>ANN</i>	<i>SO₄</i>	-0.55	<i>-0.61</i>	0.10	<i>0.60</i>
	<i>CO₂</i>			0.63	<i>0.01</i>

Table 3: Global and regional “between-experiment” spatial pattern correlations. Results are for seasonally- and annually-averaged near-surface temperature signals in the three TP perturbation experiments, with forcing due to sulfate aerosols, CO₂, and a combination of sulfate aerosols and CO₂. The perturbations approximately correspond to present-day atmospheric CO₂ levels and anthropogenic sulfur emissions. The signals are the mean changes (computed with 20-year samples) relative to the average over years 11-30 of the TP control integration with no anthropogenic sulfate aerosols and pre-industrial atmospheric CO₂ levels. The non-italicized numbers are the (centered) pattern correlations obtained using the full spatial fields, and the italicized numbers are the correlations over the area defined by the observed data mask for $t_0 = 1954$ (see Figure A2, panel b).

SEASON	EXPT	$R(t)$ vs $\Delta\hat{D}(t)$	$C(t)$ vs $\Delta\hat{D}(t)$
DJF	SO_4	-0.29	-0.98
	CO_2	0.15	0.99
	$SO_4 + CO_2$	-0.38	0.93
MAM	SO_4	-0.26	-0.99
	CO_2	0.20	1.00
	$SO_4 + CO_2$	-0.19	0.96
JJA	SO_4	-0.11	-0.95
	CO_2	0.39	0.99
	$SO_4 + CO_2$	0.06	0.63
SON	SO_4	-0.05	-0.91
	CO_2	-0.05	0.98
	$SO_4 + CO_2$	0.01	0.82
ANN	SO_4	-0.06	-0.99
	CO_2	0.13	1.00
	$SO_4 + CO_2$	-0.01	0.91

Table 4: Correlations between $C(t)$ and $R(t)$ time series and $\Delta\hat{D}(t)$, the changes in observed area-averaged near-surface temperature. Seasonal and annual $C(t)$ and $R(t)$ time series are for the three TP perturbation experiments (see Figures 6-8). Values of $\Delta\hat{D}(t)$ were computed using the observed data mask for $t_0 = 1954$ (see Figure A2, panel b). For $C(t)$, the changes for all three perturbation experiments correlate strongly with $\Delta\hat{D}(t)$, which points to a dominant influence of $\Delta\hat{D}(t)$ in the behavior of $C(t)$.

<i>EXPT</i>	<i>SEASON</i>	<i>b/a</i>	$\overline{s_D^2}$	$\overline{s_M^2}$
<i>SO₄</i>	<i>DJF</i>	1.86	0.66	0.98
	<i>MAM</i>	4.05	0.37	0.23
	<i>JJA</i>	5.00	0.21	0.27
	<i>SON</i>	2.94	0.25	1.03
	<i>ANN</i>	4.95	0.20	0.38
<i>CO₂</i>	<i>DJF</i>	2.29	0.66	0.71
	<i>MAM</i>	5.09	0.37	0.18
	<i>JJA</i>	6.68	0.21	0.18
	<i>SON</i>	4.82	0.25	0.33
	<i>ANN</i>	7.31	0.20	0.19
<i>CO₂+SO₄</i>	<i>DJF</i>	1.24	0.66	0.26
	<i>MAM</i>	1.77	0.37	0.24
	<i>JJA</i>	1.25	0.21	0.41
	<i>SON</i>	1.27	0.25	0.33
	<i>ANN</i>	2.23	0.20	0.20

Table 5: Decomposition of $C(t)$. As shown by Santer et al. (1993), $C(t)$ can be decomposed into a term measuring the strength of the pattern similarity and a term related to the observed time-varying spatial mean (see Section 5.1.4). The coefficients a and b give the relative strengths of these two terms. Values of the ratio b/a much larger than 1.0 indicate that $C(t)$ is dominated by observed changes in global-mean temperature, and provides limited pattern similarity information. This is the case for the experiments with forcing by CO_2 only and sulfate aerosols only. Also shown are the (time-averaged) spatial variance terms $\overline{s_D^2}$ and $\overline{s_M^2}$, which were computed using the observed data mask (see Figure A2, panel b).

		TREND LENGTH (YEARS)									
		10		20		30		40		50	
SEASON											
$R(t)$	<i>DJF</i>	0.344	<i>0.425</i>	0.381	<i>0.456</i>	0.091	<i>0.184</i>	0.676	<i>0.672</i>	0.737	<i>0.771</i>
$R(t)$	<i>MAM</i>	0.577	<i>0.545</i>	0.780	<i>0.749</i>	0.329	<i>0.373</i>	0.353	<i>0.350</i>	0.254	<i>0.233</i>
$R(t)$	<i>JJA</i>	0.815	<i>0.791</i>	0.380	<i>0.330</i>	0.469	<i>0.394</i>	0.557	<i>0.515</i>	0.215	<i>0.091</i>
$R(t)$	<i>SON</i>	0.535	<i>0.556</i>	0.494	<i>0.475</i>	0.358	<i>0.335</i>	0.435	<i>0.376</i>	0.887	<i>0.852</i>
$R(t)$	<i>ANN</i>	0.283	<i>0.332</i>	0.325	<i>0.358</i>	0.059	<i>0.086</i>	0.281	<i>0.201</i>	0.417	<i>0.335</i>
$C(t)$	<i>DJF</i>	0.014	<i>0.053</i>	0.000	<i>0.000</i>	0.000	<i>0.000</i>	0.000	<i>0.000</i>	0.000	<i>0.000</i>
$C(t)$	<i>MAM</i>	0.000	<i>0.006</i>	0.000	<i>0.000</i>	0.000	<i>0.000</i>	0.000	<i>0.000</i>	0.000	<i>0.000</i>
$C(t)$	<i>JJA</i>	0.240	<i>0.354</i>	0.000	<i>0.000</i>	0.000	<i>0.000</i>	0.000	<i>0.000</i>	0.000	<i>0.000</i>
$C(t)$	<i>SON</i>	0.601	<i>0.573</i>	0.000	<i>0.000</i>	0.000	<i>0.001</i>	0.000	<i>0.000</i>	0.026	<i>0.040</i>
$C(t)$	<i>ANN</i>	0.029	<i>0.095</i>	0.000	<i>0.000</i>	0.000	<i>0.000</i>	0.000	<i>0.000</i>	0.000	<i>0.000</i>
Sample size (m)		579	979	569	969	559	959	549	949	539	939

Table 6: Significance levels (p -values) for seasonal and annual near-surface temperature signals from the TP experiment with present-day CO_2 forcing. The signals are the linear trends, β_s , for the most recent 10, 20, ..., 50 years of the $R(t)$ and $C(t)$ time series shown in Figure 6 (i.e., the trends over years 1984-93, 1974-93, ..., 1944-93). To determine the significance of β_s , we require information on the behavior of the centered and uncentered pattern correlation statistics in the absence of external forcing. This information was obtained by correlating the seasonal and annual temperature-change signals from the TP CO_2 -only experiment with temperature anomalies from the 600-year HAMCTL and 1,000-year GFDLCTL integrations, each of which was performed with a fully-coupled A/OGCM. The resulting $R_N(t)$ and $C_N(t)$ time series for each coupled model control run were then used to generate sampling distributions of "unforced" linear trends, β_N , for different seasons and trend lengths. The p -values were then computed by comparing β_s with the sampling distribution β_N for the appropriate season, trend length, statistic type, and coupled model control run (see Section 6.1). Non-italicized numbers are the p -values based on β_N estimates from HAMCTL; results for GFDLCTL are shown in italics. Signal trends significant at the 5% level or better are in shaded boxes. The p -values were computed using the maximum overlap (i.e., by all but one year) in 10- to 50-year chunks of the $R_N(t)$ and $C_N(t)$ time series. The different sample sizes for the two coupled models are given in the final row.

		TREND LENGTH (YEARS)									
		10		20		30		40		50	
SEASON											
<i>R(t)</i>	<i>DJF</i>	0.960	0.916	0.884	0.824	0.952	0.872	0.749	0.653	0.659	0.558
<i>R(t)</i>	<i>MAM</i>	0.896	0.853	0.738	0.676	0.621	0.556	0.840	0.775	0.191	0.204
<i>R(t)</i>	<i>JJA</i>	0.756	0.682	0.641	0.577	0.517	0.464	0.257	0.282	0.033	0.026
<i>R(t)</i>	<i>SON</i>	0.812	0.729	0.612	0.513	0.497	0.410	0.271	0.248	0.011	0.009
<i>R(t)</i>	<i>ANN</i>	0.950	0.914	0.791	0.718	0.717	0.636	0.397	0.339	0.082	0.040
<i>C(t)</i>	<i>DJF</i>	0.520	0.508	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>C(t)</i>	<i>MAM</i>	0.124	0.178	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>C(t)</i>	<i>JJA</i>	0.382	0.431	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>C(t)</i>	<i>SON</i>	0.758	0.689	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>C(t)</i>	<i>ANN</i>	0.484	0.496	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>Sample size (m)</i>		579	979	569	969	559	959	549	949	539	939

Table 7: As for Table 6, but for seasonal and annual near-surface temperature signals from the TP experiment with combined sulfate aerosol / CO₂ forcing.

		<i>TREND LENGTH (YEARS)</i>									
		<i>10</i>		<i>20</i>		<i>30</i>		<i>40</i>		<i>50</i>	
<i>SEASON</i>											
<i>R(t)</i>	<i>DJF</i>	0.777	0.720	0.909	0.828	0.980	0.959	0.719	0.631	0.547	0.407
<i>R(t)</i>	<i>MAM</i>	0.964	0.924	0.640	0.610	0.846	0.825	0.989	0.964	0.488	0.386
<i>R(t)</i>	<i>JJA</i>	0.520	0.549	0.712	0.711	0.685	0.716	0.536	0.597	0.193	0.239
<i>R(t)</i>	<i>SON</i>	0.630	0.625	0.536	0.506	0.589	0.552	0.368	0.376	0.000	0.006
<i>R(t)</i>	<i>ANN</i>	0.946	0.908	0.784	0.797	0.986	0.972	0.796	0.778	0.184	0.146
<i>C(t)</i>	<i>DJF</i>	0.993	0.965	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
<i>C(t)</i>	<i>MAM</i>	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
<i>C(t)</i>	<i>JJA</i>	0.786	0.658	1.000	0.999	1.000	1.000	1.000	0.995	1.000	0.982
<i>C(t)</i>	<i>SON</i>	0.501	0.477	0.993	0.986	1.000	0.977	0.980	0.908	0.586	0.554
<i>C(t)</i>	<i>ANN</i>	0.983	0.930	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
<i>Sample size (m)</i>		581	979	571	969	561	959	551	949	541	939

Table 8: As for Table 6, but for seasonal and annual near-surface temperature signals from the TP experiment with present-day sulfate aerosol forcing.