

Towards the detection and attribution of an anthropogenic effect on climate

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

¹ Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

² Global Climate Research Division, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

³ National Center for Atmospheric Research, Boulder, CO 80307-3000, USA

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

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

Received: 10 November 1994 / Accepted: 19 July 1995

Abstract. It has been hypothesized recently that regional-scale cooling caused by anthropogenic sulfate aerosols may be partially 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 test this hypothesis. We use centered [$R(t)$] and uncentered [$C(t)$] pattern similarity statistics to compare observed time-evolving surface temperature change patterns with the model-predicted equilibrium signal patterns. We show that in most cases, the $C(t)$ statistic reduces to a measure of observed global-mean temperature changes, and is of limited use in attributing observed climate changes to a specific causal mechanism. 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 over the last 50 years, the summer (JJA) and fall (SON) observed patterns of near-surface temperature change show increasing similarity to the model-simulated response to combined sulfate aerosol/ CO_2 forcing. At least some of this increasing spatial congruence occurs in areas where the real world has cooled. To assess the significance of the most recent trends in $R(t)$ and $C(t)$, we use data from multi-century control integrations performed with two different coupled atmosphere-ocean models, which provide information on the statistical behavior of ‘unforced’ trends in the pattern correlation statistics. For the combined sulfate aerosol/ CO_2 experiment, the 50-year $R(t)$ trends for the JJA and SON signals are highly significant. 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 CO_2 -only signal are not significant in any season. $C(t)$ trends for signals from both the CO_2 -only and combined forcing experiments are highly significant in

all seasons and for all trend lengths (except for trends over the last 10 years), indicating large global-mean changes relative to the two natural variability estimates used here. 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 CO_2 /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 that we have identified is distinctly different from internally generated natural variability noise. The fact that we have been able to detect the detailed spatial signature in response to combined CO_2 and sulfate aerosol forcing, but not in response to CO_2 forcing alone, suggests that some of the regional-scale background noise (against which we were trying to detect a CO_2 -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 “fingerprint” strategy introduced by Madden and Ramanathan (1980; see also MacCracken and Moses 1982) have used some form of pattern correspondence 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 CO_2 -doubling experiment, or towards the end of an experiment with a time-dependent CO_2 increase.

The strategy in this method is to search for a long-term, positive trend in the pattern correspondence statistic, which would indicate an increasing expression of the CO₂ 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 centered 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 CO₂ 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 CO₂ 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. (Unless the change in the global mean is so large that it cannot be explained by non-anthropogenic forcing mechanisms and/or natural variability.) The studies cited above suggest that 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 sub-global-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 that employed a centered statistic to search observed records of near-surface air temperature for model-predicted CO₂ 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 CO₂ signal pattern, or masking of sub-global-scale features of the

CO₂ signal by low-frequency natural variability and/or other forcing factors (anthropogenic sulfate aerosols, volcanic aerosols, solar variability, etc.).

Model experiments recently performed by Taylor and Penner (1994; henceforth TP) 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 mainly from the SO₂ emitted by fossil fuel combustion. 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 mixed-layer ocean and a tropospheric chemistry model to investigate the climate response to the individual and combined forcing by anthropogenic sulfate aerosols and CO₂. The results from these experiments 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 atmosphere-ocean GCMs (CGCMs; Stouffer et al. 1989; Cubasch et al. 1992; Meehl et al. 1993).

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. (1995) and Mitchell et al. (1995a,b). These investigations also considered the temperature response to a combination of CO₂ and anthropogenic sulfate aerosol forcing, but the direct radiative effect of the aerosol was parameterized by changing the surface albedo. Neither study allowed changes in climate to influence the aerosol distribution, whereas in the TP study two-way interaction is possible between climate and aerosols.

To date, only two previous attempts have 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 (Santer et al. 1995a; Mitchell et al. 1995b). Related 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 are likely to have been large (Wigley et al. 1992; Engardt and Rodhe 1993; Hunter et al. 1993; Karl et al. 1995) 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 investigate 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 structure of this paper is as follows. Section 2 provides a brief description of the TP numerical experiments. In Sect. 3 we compare the surface temperature responses in the three TP perturbation experiments using maps of the geographical distribution of temperature change, pattern correlations, and univariate t-tests. In Sect. 4, we define centered and uncentered pattern similarity statistics 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 signals from the three TP perturbation experiments. The question of whether the trends in our pattern similarity measures are significant is addressed in Sect. 6. A short summary and conclusions are given in Sect. 7.

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 AGCM 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 × 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 CO₂ levels (270 ppmv; the standard IPCC value is 278 ppmv, Enting 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; Benkowitz 1982) and near-pre-industrial CO₂ concentrations; a CO₂-only experiment (C) with no sulfur emissions and nominal present-day CO₂ levels (345 ppmv; c.f. the 1990 value of 354 ppmv given by Enting et al. 1994); and a combined forcing experiment with near-present-day sulfur emissions and CO₂. 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 a spin-up period of at least 10 years).

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). For comparing patterns, we computed centered spatial correlations, here denoted by R (see Santer et al. 1993). 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 in DJF (ca. -7°C) and to the east of the Weddell Sea in JJA (-6°C).

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. Figs. 1, 2, and Fig. 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. (1995) and Mitchell et al. (1995b).

There are substantial uncertainties associated with the forcing and response patterns simulated in such studies, and hence with the forcing versus response pattern correlations presented above. In both the Roeckner et al. (1995) and TP investigations, these uncertainties are due in part to inadequacies in the treatment of sea-ice and neglect of full ocean dynamics. The latter does not apply to the study by Mitchell et al. (1995b), which used a full CGCM and still found the patterns of temperature response and forcing due to the direct effect of sulfate aerosols to be dissimilar.

The results of all three investigations illustrate the potential importance of the atmospheric general circulation in ‘modulating’ the response to regionally-localized forcing. Detection studies that use as their signal the *global* sulfate aerosol distribution or aerosol forcing predicted by a sulfur chemistry model (e.g., Wigley et al. 1992; Engardt and Rodhe 1993), or the global pattern of observed changes in sulfur emissions (Karl et al. 1995) may therefore be using sub-optimal signals. On a regional basis, however, it is likely that there is a close relationship between emissions, forcing and response in areas with high sulfur emissions. This would explain Karl et al.’s (1995) finding of a significant in-

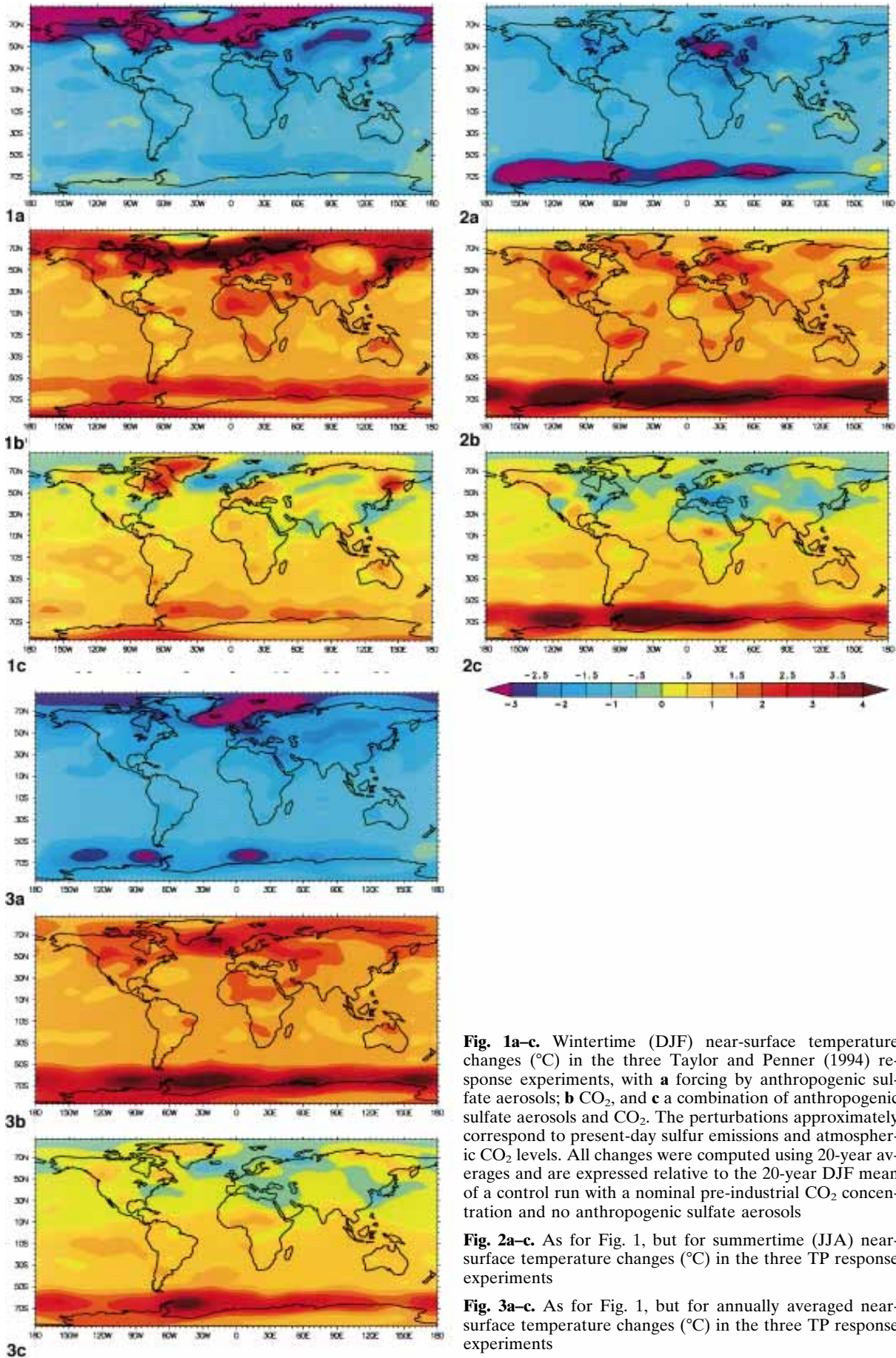


Table 1. Global spatial pattern correlations between forcing and response fields

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 ₄	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

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 Fig. 4). The response fields are the time-averaged near-surface temperature changes in Figs. 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)

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

Season	Experiment	$\alpha=0.01$	$\alpha=0.05$
DJF	SO ₄	86.5	93.0
	CO ₂	91.8	95.8
	SO ₄ +CO ₂	56.1	65.3
MAM	SO ₄	89.4	93.7
	CO ₂	89.6	95.4
	SO ₄ +CO ₂	55.2	65.7
JJA	CO ₄	88.3	93.4
	CO ₂	90.0	94.4
	SO ₄ +CO ₂	55.5	64.5
SON	SO ₄	87.2	92.6
	CO ₂	95.8	98.3
	SO ₄ +CO ₂	54.1	65.0
ANN	SO ₄	98.8	99.2
	CO ₂	100.0	100.0
	SO ₄ +CO ₂	65.6	71.1

All tests were 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

verse relationship between changes in sulfur emissions and changes in surface temperature in zones of high emissions (50°N–55°N) and at times of high insolation.

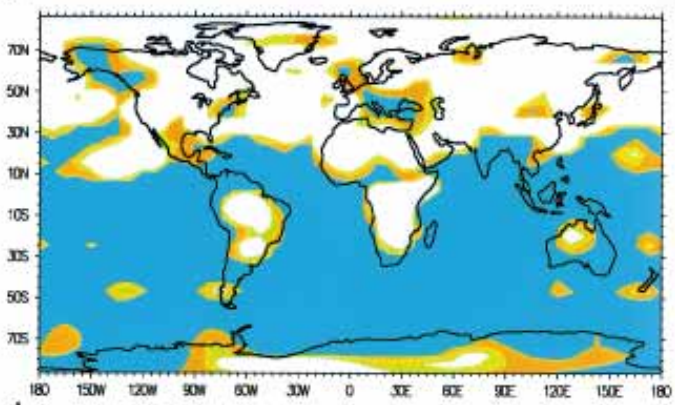
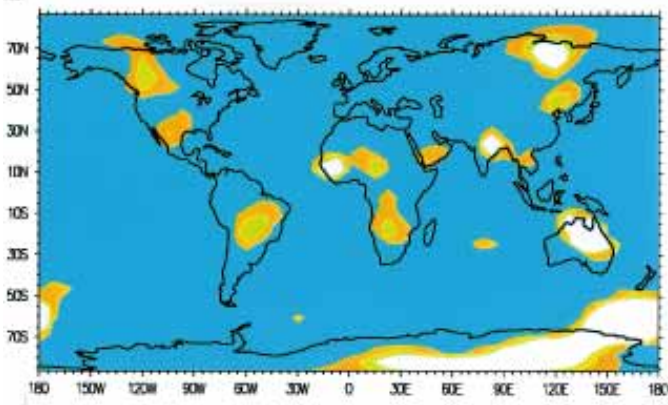
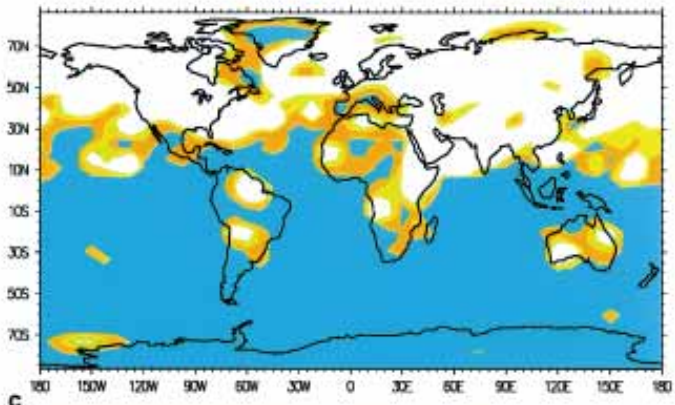
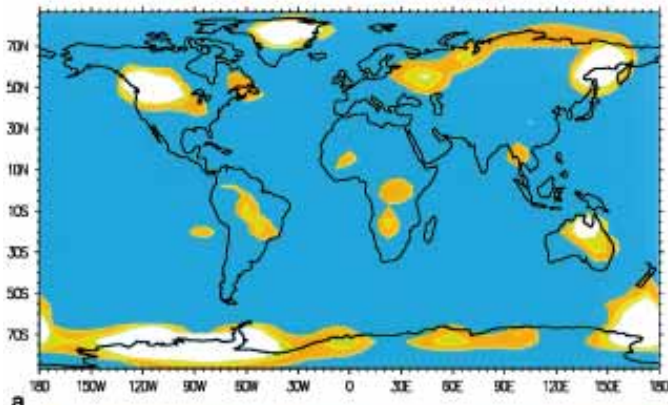
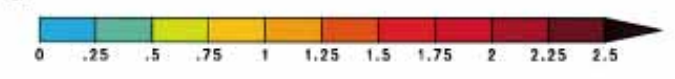
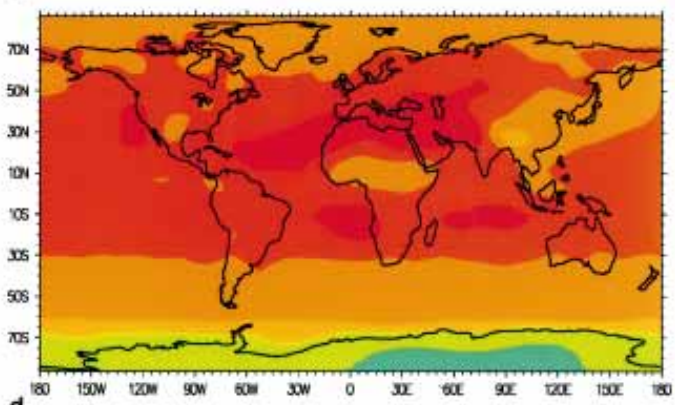
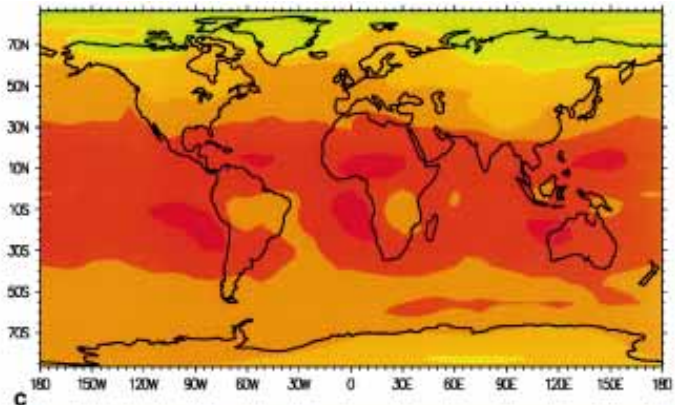
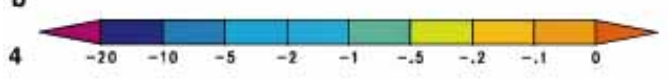
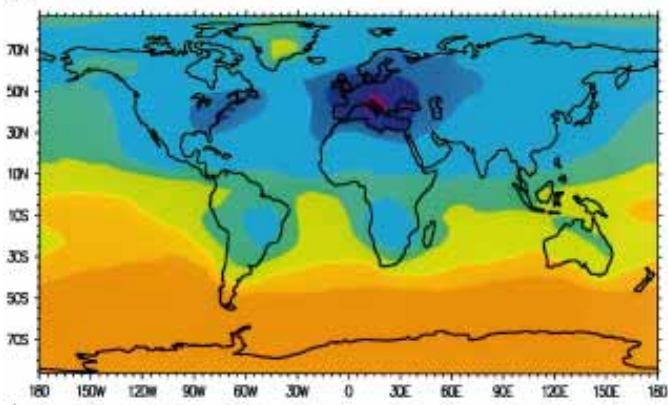
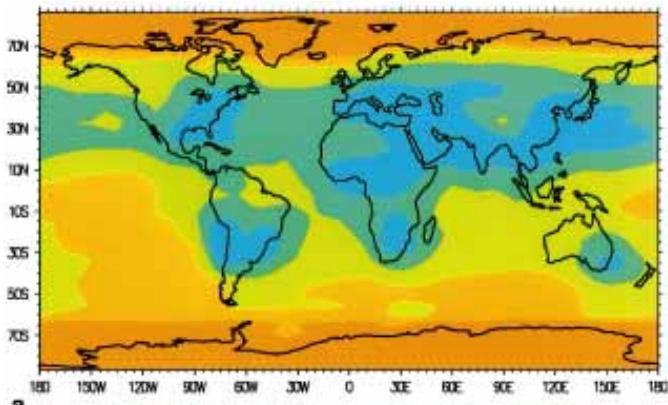
Note that the maximum Northern Hemisphere cooling occurs in DJF, even though the maximum forcing is in JJA and the DJF forcing is relatively small (c.f. Figs. 1a, 2a, and Fig. 4). In JJA, the maximum response is in the Southern Hemisphere, even though the forcing is largely in the Northern Hemisphere. 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 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-test reveals that the changes in mean state are highly significant in the sulfate-only experiment (Table 2 and Fig. 5). (All t-tests were performed using 20-year samples from the control run and the perturbation experiments). Of the total number of grid-point tests performed 86.5–98.8% 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) 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 (Figs. 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 commonly used in step-function CO₂-doubling 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 reduc-



tion 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 CGCMs (Cubasch et al. 1992; Santer et al. 1994).

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. Figs. 1b, 2b, 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). It should be noted that in TP, the climate sensitivity of the model for a doubling of CO₂ was given as 6.4°C, a result that was based on 10-year samples only. The results 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. This yielded 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 (Figs. 1c, 2c, 3c). Temperature decreases are restricted largely to

Table 3. Global and regional ‘between-experiment’ spatial pattern correlations

Season	Experiment	CO ₂	SO ₄ + CO ₂		
DJF	SO ₄	-0.64	-0.73	0.46	0.48
	CO ₂			0.01	-0.18
MAM	SO ₄	-0.53	-0.31	-0.26	0.49
	CO ₂			0.79	0.29
JJA	SO ₄	-0.59	-0.30	-0.26	0.58
	CO ₂			0.79	0.32
SON	SO ₄	-0.38	-0.40	0.39	0.59
	CO ₂			0.52	0.20
ANN	SO ₄	-0.55	-0.61	0.10	0.60
	CO ₂			0.63	0.01

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 Fig. A2, panel b)

the Northern Hemisphere, which is where most of the radiative forcing associated with sulfate aerosols occurs (Fig. 4a, b). The largest decreases are over the Norwegian Sea in DJF (ca. -2°C) and over southeastern 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 (Fig. 1a, b) are considerably reduced in the SC integration (Fig. 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. Figs. 1a, b, 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 due to compensating warming and cooling responses over large areas

Fig. 4a–d. Seasonally-averaged radiative forcing due to sulfate aerosols and CO₂ in the TP sulfate-only and CO₂-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 **a** for winter (DJF); and **b** summer (JJA). Changes in the net long-wave flux at the tropopause due to an increase in atmospheric CO₂ from 270 ppmv to 345 ppmv are also given for **c** winter and **d** summer

Fig. 5a–d. Univariate t-test results for seasonally-averaged near-surface temperature changes in the TP response experiments. Results are for **a** DJF and **b** JJA changes in the sulfate-only experiment and in the experiment with combined CO₂/sulfate aerosol forcing (**c**, **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 CO₂-only experiment (not shown) are highly significant at virtually all grid-points

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 (Fig. 5).

The weak response over large areas of the Northern Hemisphere, as indicated by a lack of statistical significance, 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 Sect. 5.1, which presents quantitative measures of observed versus simulated pattern similarity.

4 Pattern similarity statistics

In this section we introduce the pattern similarity statistics, $R(t)$ and $C(t)$, which we subsequently use for comparing model and observed spatial patterns of temperature change. 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 the 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 dependence on season, since the TP integrations are equilibrium response experiments with no interannual changes in the forcing. The rationale for the use of equilibrium signal patterns (as opposed to transient signals) is discussed in detail in Santer et al. (1995a).

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. They contain high-frequency variability components associated with El Niño behavior, volcanoes, etc. For the purposes of GHG-detection studies, these components constitute undesirable noise; data were filtered in the following way to reduce this noise.

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 then define the (smoothed) anomaly at time t relative to $\overline{D}_1(x)$ by

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

where

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

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 $W_{\text{crit}}=0.6$, which allows us to calculate filtered means for 1993. Further details of the filtering procedure are given in Santer et al. (1995a).

The set of time-evolving monthly-mean anomaly fields, $\Delta D(x, t)$, was then used to compute seasonal and annual averages. In order to avoid large, spatially non-random increases in coverage from the beginning of the century to the present, we stipulated that the grid-point coverage for $\overline{D}_2(x, t)$ must be a subset of the coverage for $\overline{D}_1(x)$ (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, $\Delta D(x, t)$.

Following Santer et al. (1993), we use two different types of measure to compare the spatial fields $\Delta M(x)$ and $\Delta D(x, t)$. 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

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

is the spatial variance (with s_M^2 defined similarly), and the $\widehat{\quad}$ in Eqs. (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 uncentered 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 Eq. (7). In contrast, a trend in $R(t)$ with time may 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, $\Delta D(t)$, which is an important part of the signal in most situations, while $R(t)$ focuses on the pattern of change. We have shown previously (Santer et al. 1993, 1995a) that $C(t)$ can be decomposed into $R(t)$ and $\widehat{\Delta D}(t)$ 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 $\widehat{\Delta D}(t)$ component, so time series of $C(t)$ look very similar to those of $\widehat{\Delta D}(t)$ (see e.g., Santer et al. 1993, Fig. 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 focuses on the pattern of spatial anomalies about the mean state, may provide the information required to discriminate between forcing mechanisms with different pattern signatures.

In the following, our emphasis is on attribution; hence we prefer to use $R(t)$. We also compute $C(t)$, since this enables us to determine the significance of global-mean changes and to compare our results with other investigations that have used uncentered statistics (e.g., Hegerl et al. 1994).

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 show 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 (Fig. 6). If CO₂ forcing were the dominant influence on climate, i.e., if other external forcings and natural internal variability were relatively small 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 as the signal became progressively more pronounced in the observed data.

$C(t)$ shows no evidence of large, positive trends that are sustained for 40 to 50 years or longer (Fig. 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). Changes in $C(t)$ parallel those in global-mean temperature over the entire observed record (Jones et al. 1991; Jones and Briffa 1992). Note also that in all seasons except DJF, $C(t)$ time series 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 in agreement with the theoretical and empirical results of Santer et al. (1993, 1995a). As shown in Table 4, $C(t)$ changes for all three perturbation experiments correlate strongly with $\widehat{\Delta D}(t)$, although the correlations are smallest for the combined forcing case. This points to a dominant influence of $\widehat{\Delta D}(t)$ on the behavior of $C(t)$, as discussed further in Santer et al. (1995a).

$R(t)$ behaves quite differently from $C(t)$ (see Fig. 6 and Table 4). 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. The

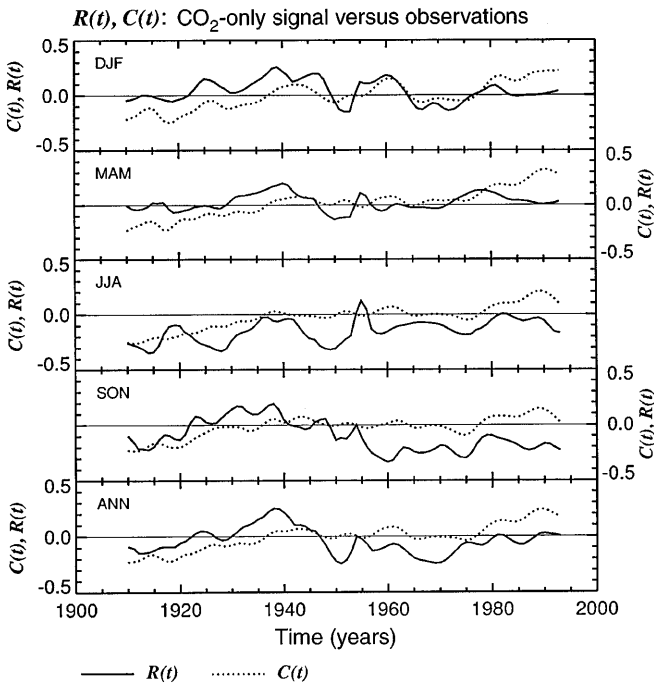


Fig. 6. Centered $R(t)$ and uncentered $C(t)$ pattern correlations between model and observed near-surface temperature changes. Model changes in seasonally and annually averaged temperature are taken from the the TP CO_2 -only experiment. For each season, one time-independent spatial pattern characterizes the response to forcing by present-day atmospheric CO_2 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

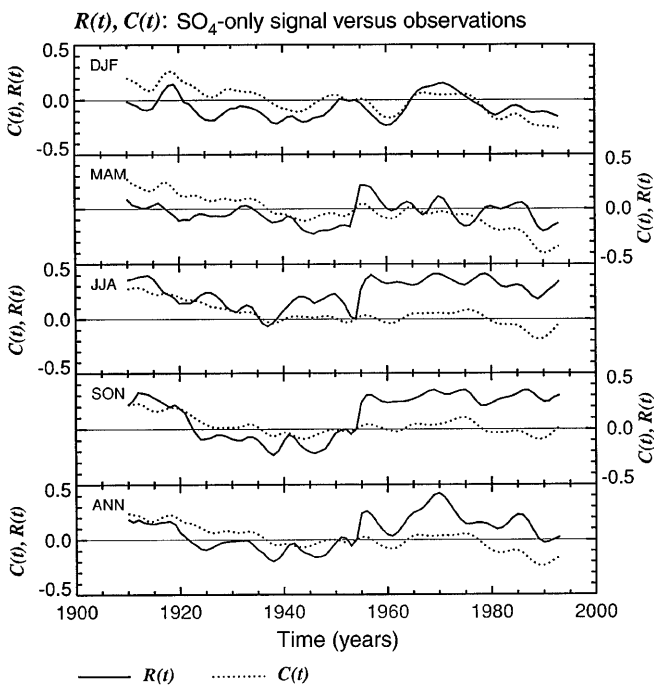


Fig. 7. As for Fig. 6, but for near-surface temperature changes from the TP sulfate-only experiment

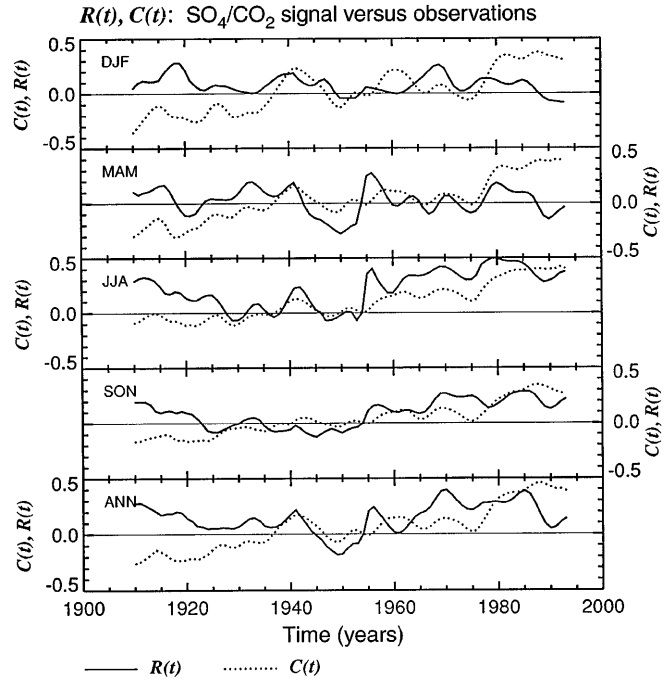


Fig. 8. As for Fig. 6, but for near-surface temperature changes from the TP experiment with combined CO_2 and sulfate aerosol forcing

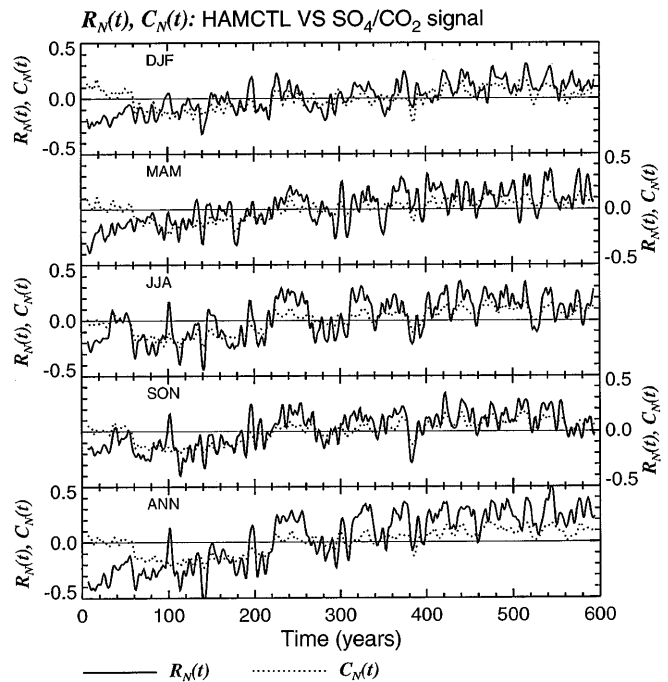


Fig. 9. Behavior of pattern correlation statistics in the absence of external forcing. $R_N(t)$ and $C_N(t)$ time series were computed using (filtered) time-dependent patterns of near-surface temperature change from a 600-year control integration (HAMCTL) with a CGCM and fixed seasonal- and annual temperature-change patterns from the TP experiment with combined CO_2 /sulfate aerosol forcing

Table 4. Correlations between $C(t)$ and $R(t)$ and $\widehat{\Delta D}(t)$, the changes in observed area-averaged near-surface temperature

Season	Experiment	$R(t)$ vs. $\widehat{\Delta D}(t)$	$C(t)$ vs. $\widehat{\Delta D}(t)$
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

Seasonal and annual $C(t)$ and $R(t)$ time series are for the three TP perturbation experiments (see Figs. 6–8). Values of $\widehat{\Delta D}(t)$ were computed using the observed data mask for $t_0=1954$ (see Fig. A2, panel b). For $C(t)$, the changes for all three perturbation experiments correlate strongly with $\widehat{\Delta D}(t)$, which points to a dominant influence of $\widehat{\Delta D}(t)$ in the behavior of $C(t)$

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. Figs. 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 pattern similarity statistic, 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 CO₂-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 quite different from the corresponding time series for the CO₂-only experiment (c.f. Figs. 6 and 8). This is most pronounced in JJA and SON: $R(t)$ now shows a large positive trend that is sustained over the last 40–60 years of the observed record. The positive trends in JJA and SON are similar to those for the sulfate-only case (Fig. 7). This implies 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 CO₂ 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 1940–1970 period, when $R(t)$ shows little trend for the CO₂-only signal (Fig. 6).

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. 1995) with the TP signal patterns from the combined forcing experiment. For JJA, such a comparison shows that the observed cooling over south-eastern Europe, the east coast of the U.S., and the Tibetan/Mongolian Plateau is congruent with the TP temperature change pattern (c.f. our Fig. 2c with Fig. 2f in Karl et al. 1995). This must contribute towards the large 50-year trends in $R(t)$.

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 CO₂ and sulfate aerosol forcing show large, positive linear trends over the last 40–60 years. 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: i.e., trends due solely to the effects of natural variability on decadal- to century-time scales. We cannot obtain this information easily from the observed data in view of the difficulties involved in 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 the model world, however, we can examine output from experiments with no external forcing in order to estimate the magnitude and spatial properties of internally-generated natural variability noise. In this study we therefore employ model-generated noise data to assess trend significance. We stress that the model noise is due only to internal variability of the coupled atmosphere-ocean system, and does not incorporate variability related to changes in natural external forcings (e.g., in solar output or the volcanic dust loading of the atmosphere).

The 20-year TP control integration is clearly too short for estimating natural variability noise characteristics on time scales appropriate to a slowly-evolving anthropogenic signal. We therefore used surface temperature results from two recent multi-centennial control integrations performed with fully-coupled atmosphere-ocean GCMs to determine the statistical behavior 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

CGCM (Hegerl et al. 1994) (the ECHAM-1 T21 resolution AGCM coupled to the large-scale geostrophic OGCM), and a 1000-year integration of the GFDL coupled model (Delworth et al. 1993; Stouffer et al. 1994) (the GFDL R15 AGCM coupled to the modular ocean model). 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 study. However, since data from these integrations are used in the significance testing analysis, we provide an overview of model variability differences in Appendix B.

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 ‘unforced’ trends in these statistics. The method we use is similar to that employed by Santer et al. (1995b).

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

$$\Delta N(x, t) = \overline{N}_2(x, t) - \overline{N}_1(x) \quad (8)$$

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 Eq. (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 $t_0=1954$ (see Fig. A2) were excluded from the analysis. Unlike the situation when dealing with observed data (see Appendix A), there are no coverage changes as a function of time: once the (filtered) observed data mask for $t_0=1954$ has been used to exclude model data points from the analysis, this mask does not change with time. We then substituted $\Delta N(x, t)$, $\overline{\Delta N}$, etc. for the corresponding observed terms in Eqs. (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 (Fig. 9). Both sets of time series have considerable variability on 10- to

20-year time scales. For $C(t)$, the observed changes over the last 40- to 50 years (≈ 0.3 – 0.4 in the case of the CO_2 -only signal; see Fig. 6) are noticeably larger than $C_N(t)$ changes (≈ 0.2). For $R(t)$, however, it is more difficult to evaluate whether the 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. This yields a distribution of slope parameters, $\beta_N(i)$, $i=1, \dots, 5$, for each statistic, season, response experiment, and natural variability noise integration. These distributions are the yardsticks against which we 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 Figs. 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 Fig. 10). 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.

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 $\beta_N(i) \geq \beta_S(i)$. Note 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. The probability of obtaining the signal trend by chance based on the natural variability manifested in HAMCTL or GFDLCTL is then simply

$$p_i = k_i / m_i \quad (9)$$

where m_i is the sample size for $\beta_N(i)$. Note that the number of *independent* linear trend samples in HAMCTL and GFDLCTL may differ from m_i due to the long decorrelation time of near-surface temperature, both in models and in the observations (see Hegerl et al. 1994). The most satisfactory way of increasing the number of independent samples would be by extending the two control runs used here for thousands of (simulated!) years. Such lengthy CGCM integrations may become feasible in the next few years, but were not available for the purposes of this investigation.

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; Santer et al. 1995b), although this does not lead to a proportionate increase in the number of independent samples in the distribution. We computed p -values using both non-overlapping chunks and chunks that had the maximum

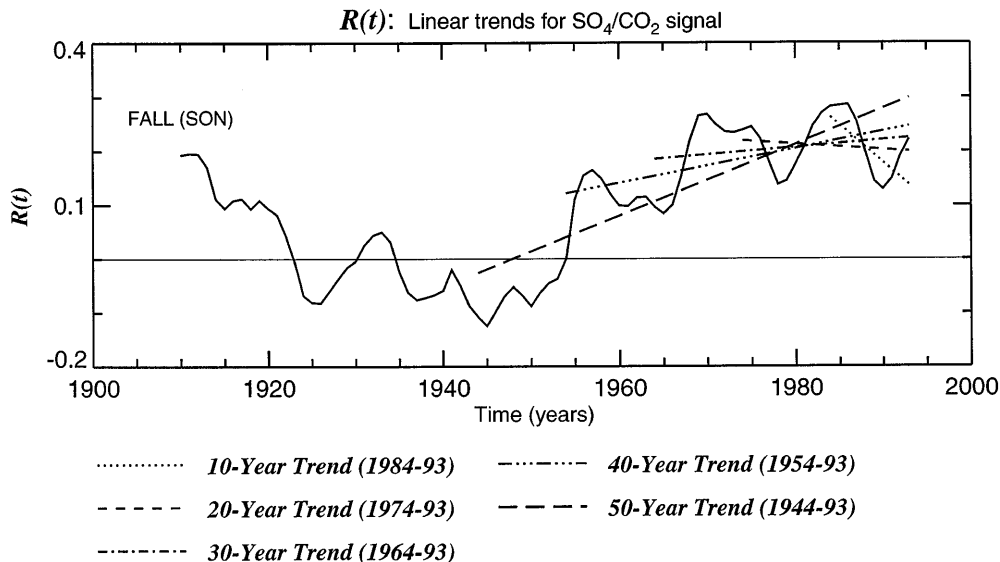


Fig. 10. Linear trends for the final 10–50 years of the $R(t)$ time series for the SON near-surface temperature signal from the TP experiment with combined CO_2 /sulfate aerosol forcing. These are the signal trends $\beta_S(i)$, whose significance we wish to determine

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

CO_2 -only signal. The p -values for the near-surface temperature signal from the TP CO_2 -only experiment

are given in Table 5 as a function of trend length, season, statistic, and natural variability noise experiment. Based on noise levels from either of the natural variability experiments, 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 CO_2 -only signal (Table 4), our results imply that the most recent 20- to 50-year trends in observed global-mean temperature are large relative to the variability of ‘unforced’ linear trends in HAMCTL and GFDLCTL.

Numerous other studies have demonstrated that the

Table 5. Significance levels (p -values) for seasonal and annual near-surface temperature signals from the TP experiment with present-day CO_2 forcing

		Trend length (years)									
		10		20		30		40		50	
Season											
$R(t)$	DJF	0.34	<i>0.43</i>	0.38	<i>0.46</i>	0.09	<i>0.18</i>	0.68	<i>0.67</i>	0.74	<i>0.77</i>
$R(t)$	MAM	0.58	<i>0.55</i>	0.78	<i>0.75</i>	0.33	<i>0.37</i>	0.35	<i>0.35</i>	0.25	<i>0.23</i>
$R(t)$	JJA	0.82	<i>0.79</i>	0.38	<i>0.33</i>	0.47	<i>0.39</i>	0.56	<i>0.52</i>	0.22	<i>0.09</i>
$R(t)$	SON	0.54	<i>0.56</i>	0.49	<i>0.48</i>	0.36	<i>0.34</i>	0.44	<i>0.38</i>	0.89	<i>0.85</i>
$R(t)$	ANN	0.28	<i>0.33</i>	0.33	<i>0.36</i>	0.06	<i>0.09</i>	0.28	<i>0.20</i>	0.42	<i>0.34</i>
$C(t)$	DJF	0.01	<i>0.05</i>	0.00	<i>0.00</i>	0.00	<i>0.00</i>	0.00	<i>0.00</i>	0.00	<i>0.00</i>
$C(t)$	MAM	0.00	<i>0.01</i>	0.00	<i>0.00</i>	0.00	<i>0.00</i>	0.00	<i>0.00</i>	0.00	<i>0.00</i>
$C(t)$	JJA	0.24	<i>0.35</i>	0.00	<i>0.00</i>	0.00	<i>0.00</i>	0.00	<i>0.00</i>	0.00	<i>0.00</i>
$C(t)$	SON	0.60	<i>0.57</i>	0.00	<i>0.00</i>	0.00	<i>0.00</i>	0.00	<i>0.00</i>	0.00	<i>0.00</i>
$C(t)$	ANN	0.03	<i>0.10</i>	0.00	<i>0.00</i>	0.00	<i>0.00</i>	0.00	<i>0.00</i>	0.00	<i>0.00</i>
Sample size (m)		579	979	569	969	559	959	549	949	539	939

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 Fig. 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 1000-year GFDLCTL integrations, each of which was performed with a CGCM. 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 Sect. 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 bold type. The p -values were computed using overlapping 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

Table 6. As for Table 5, but for seasonal and annual near-surface temperature signals from the TP experiment with combined sulfate aerosol/CO₂ forcing

Season	Trend length (years)									
	10	20	30	40	50	10	20	30	40	50
$R(t)$ DJF	0.96	0.92	0.88	0.82	0.95	0.87	0.75	0.65	0.66	0.56
$R(t)$ MAM	0.90	0.85	0.74	0.68	0.62	0.56	0.84	0.78	0.19	0.20
$R(t)$ JJA	0.76	0.68	0.64	0.58	0.52	0.46	0.26	0.28	0.03	0.03
$R(t)$ SON	0.81	0.73	0.61	0.51	0.50	0.41	0.27	0.25	0.01	0.01
$R(t)$ ANN	0.95	0.91	0.79	0.72	0.72	0.64	0.40	0.34	0.08	0.04
$C(t)$ DJF	0.52	0.51	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$C(t)$ MAM	0.12	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$C(t)$ JJA	0.38	0.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$C(t)$ SON	0.76	0.69	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$C(t)$ ANN	0.48	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sample size (m)	579	979	569	969	559	959	549	949	539	939

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 CGCMs (Stouffer et al. 1994). Our investigation differs from such previous work in its focus on shorter period (10- to 50-year) signal trends, seasonal decomposition of the signal, and the use of noise information from two different CGCMs.

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 low- to mid-tropospheric temperature (sampled by the satellite-based microwave sounding unit, MSU) from 1979–1994. Allen et al. (1994) used detrended instrumental sea-surface temperature (SST) data to estimate the magnitude of natural variability on time scales appropriate to the length of their signal, and concluded that the 15-year trend in low-to mid-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. (1994) found that their (non-optimized) detection statistic was largely a measure of observed global-mean annually averaged temperature change for a CO₂-only signal. The most recent 20- and 30-year trends in the Hegerl et al. (1994) 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. (1994) improved upon their non-optimized results by rotating their fingerprint pattern in a direction in which the CO₂ signal could be well-represented and the natural variability noise was small (see Hasselmann 1979, 1993 for details of the optimization technique).

In contrast to the $C(t)$ results, none of our $R(t)$ results for the TP CO₂-only signal achieve significance at the 5% level (Table 5). There is little evidence, therefore, that the CO₂-only spatial pattern of temperature change (minus the global mean) 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 6). This result is similar to that obtained in the previous section for the CO₂-only signal, and arises largely through the strong correlation between $C(t)$ and $\overline{\Delta D}(t)$. It is clear that trends in $R(t)$ also contribute to the $C(t)$ results, at least in JJA and SON (see Santer et al. 1995a).

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 (GFDLCTL or HAMCTL) used to define the natural variability noise. 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 6).

For each of the TP response experiments, we have performed 50 individual significance tests of the observed $R(t)$ trends, i.e., one test for each season (plus the annual average), trend length, and natural variability experiment. A certain fraction of these tests would be expected to yield significant $R(t)$ trends due to chance alone (see Wigley and Santer 1990). For the SC signal, five $R(t)$ results were significant at the 5% level or better. The probability of obtaining this result due to chance alone is ca. 10% (this was estimated with the binomial distribution). Taken together with the result that *none* of the $R(t)$ trends for the CO₂-only signal

Table 7. As for Table 5, but for seasonal and annual near-surface temperature signals from the TP experiment with present-day sulfate aerosol forcing

Season	Trend length (years)									
	10	20	30	40	50	10	20	30	40	50
$R(t)$ DJF	0.78	0.72	0.91	0.83	0.98	0.96	0.72	0.63	0.55	0.41
$R(t)$ MAM	0.96	0.92	0.64	0.61	0.85	0.83	0.99	0.96	0.49	0.39
$R(t)$ JJA	0.52	0.55	0.71	0.71	0.69	0.72	0.54	0.60	0.19	0.24
$R(t)$ SON	0.63	0.63	0.54	0.51	0.60	0.55	0.37	0.38	0.00	0.01
$R(t)$ ANN	0.95	0.91	0.78	0.80	0.99	0.97	0.80	0.78	0.18	0.15
$C(t)$ DJF	0.99	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$C(t)$ MAM	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$C(t)$ JJA	0.79	0.66	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98
$C(t)$ SON	0.50	0.48	0.99	0.99	1.00	0.98	0.98	0.91	0.59	0.55
$C(t)$ ANN	0.98	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Sample size (m)	581	979	571	969	561	959	551	949	541	939

were significant (and with the consistently larger p -values in C than in SC for 50-year $R(t)$ trends; see Tables 5 and 6), it is highly unlikely that the $R(t)$ results in the SC and C experiments are due to chance alone.

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 7). Recall that the 50-year $R(t)$ trend in SON also achieves significance for the signal from the combined forcing experiment (Table 6). 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 5 and 7), i.e., $C(t)$ trends are significantly *negative* for most trend lengths and seasons. (Note that differences in the C versus S significance results for $C(t)$ are predominantly due to differences in $\Delta\bar{M}$, since $C(t) \approx \Delta\bar{D}(t)/\Delta\bar{M}$.)

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 CO_2 concentrations. An important question is whether the predictions by such models are consistent with the observed surface temperature record over the last century.

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 internal variability may yield similar global-mean changes.

Other investigations have attempted to address the attribution issue by defining a CO_2 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 sub-global-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 may be the neglect of other, non-greenhouse external forcings in defining the searched-for signal (e.g., changes in anthropogenic sulfate aerosols, volcanic aerosols, and solar output).

Model integrations recently carried out by Taylor and Penner (1994) represent a first attempt to simulate explicitly the temperature response to a combination of CO_2 and sulfate aerosol forcing, and formed the basis of the current study. The seasonally and annually averaged patterns of near-surface temperature change were very different in the three response experiments: a ‘sulfate-only’ experiment with near-present-day sulfur emissions (S), a ‘ CO_2 -only’ experiment with near-current atmospheric CO_2 concentrations (C), and a combined forcing experiment with near-present-day sulfur emissions and CO_2 levels (SC). While the integrations with individual forcing showed either global-scale, spatially-coherent cooling (S) or warming (C), the experiment with combined sulfate/ CO_2 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 Figs. 1–3).

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 (Jones et al. 1991). For each experiment and season, a single pattern characterizing the model signal was compared with a sequence of observed temperature change patterns for the years 1910–1993. 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 (see Sect. 4).

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. 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 behavior of $C(t)$ closely parallels changes in observed global-mean temperature.

To assess the significance of the most recent $R(t)$ and $C(t)$ trends, we used data from multi-century control integrations performed with the Hamburg and GFDL CGCMs (HAMCTL and GFDLCTL, respectively). The two control runs provide internally-consistent (but model-specific) estimates of the magnitude and patterns of surface temperature variability on long time scales.

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 the internally generated natural variability. For the signal from the TP sulfate-only experiment, the 50-year $R(t)$ trend in SON is the only result that 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 the hypothesis that we have detected 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 our assessment of trend significance. Both have attendant uncertainties.

Some of the noise uncertainties were illustrated by a preliminary comparison of the low-frequency variability of surface temperature in HAMCTL and GFDLCTL (Appendix B). This variability shows important differences in the two integrations, in terms of both global means and patterns. Some of these discrepancies are due to the different levels of climate drift in the two integrations (see Fig. B1). The primary conclusions of our study – the significance of the 50-year

$R(t)$ trends in JJA and SON for the temperature signal from the SC experiment, and the failure to find significant $R(t)$ trends for the C experiment – are not affected by these inter-model differences in variability. However, this does not guarantee that our results will remain robust when we use estimates of internally-generated natural variability from other CGCMs, as these become available.

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 in any one model to changes in resolution, physics, parameterizations, or flux correction scheme. Both control integrations considered here 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 time scale variability. Clearly, we urgently require studies that attempt to validate this variability in CGCM control runs (Barnett et al. 1995). This is a difficult task (Santer et al. 1995b), but without such work substantial uncertainties will remain regarding the reliability of the significance levels obtained here.

Similar uncertainties affect the signal patterns used in this study. 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 direct 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 arise from 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, and failure to include the effects of carbonaceous aerosols generated by biomass burning, fossil fuel combustion, and industrial processes (Penner et al. 1992, 1994b). The optical properties of sulfate aerosols are very different from those of carbonaceous aerosols. The optical properties of carbonaceous aerosols are very different from those of sulfate aerosols.

A further important signal uncertainty is our use of a model-predicted equilibrium pattern of temperature change. The use of an equilibrium signal pattern is a reasonable strategy in ‘CO₂-only’ detection studies, since there is evidence that the equilibrium and transient temperature responses to CO₂ forcing may have strong pattern similarities (see Santer et al. 1994,

1995a). This similarity is partly due to the relative stability of the pattern of CO₂-induced radiative forcing as a function of time. For sulfate aerosols, however, there have been distinct regional changes in the time-evolving pattern of forcing (e.g., changes related to declining sulfur emissions in the USA and Western Europe over the past 20 years and increasing emissions in China and India; see Karl et al. 1995). Our detection strategy does not take these changes into account. It is likely that a more complete explanation of observed changes may be achieved in detection studies based on CGCM transient experiments driven by our best estimates of observed spatio-temporal changes in GHG and sulfur emissions.

In summary, there are numerous caveats regarding the signals and natural variability noise that form the basis of this study. 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 time scales, 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.

Acknowledgements. Work at Lawrence Livermore National Laboratory was performed under the auspices of the US 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 US Department of Energy, Environmental Sciences Division, under grant DE-FG02-86ER60397. The 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 von Storch for supplying results from the 600-year HAMCTL integration. Ron Stouffer and Tom Delworth kindly supplied data from the 1000-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 $\bar{D}_1(x)$, influences the $R(t)$ and $C(t)$ results for the near-surface temperature signals from the three TP response experiments. Sensitivity of trends in $R(t)$

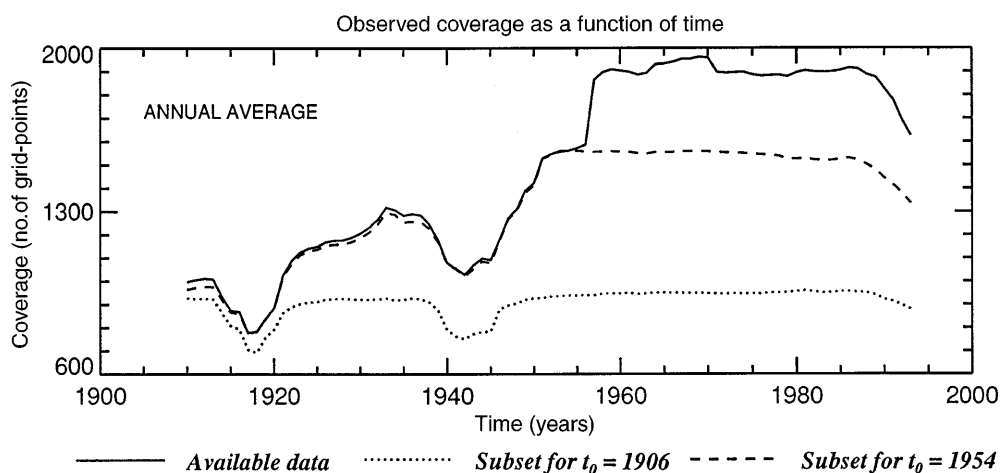


Fig. A1. Observed changes in coverage of gridded, annually averaged near-surface temperature data as a function of time. The curve “available data” shows the change in coverage for the filtered Jones et al. (1991) data if the data are used “as is” (i.e., as anomalies relative to 1950–79). The curves “ $t_0=1906$ ” 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 A). 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 1550 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

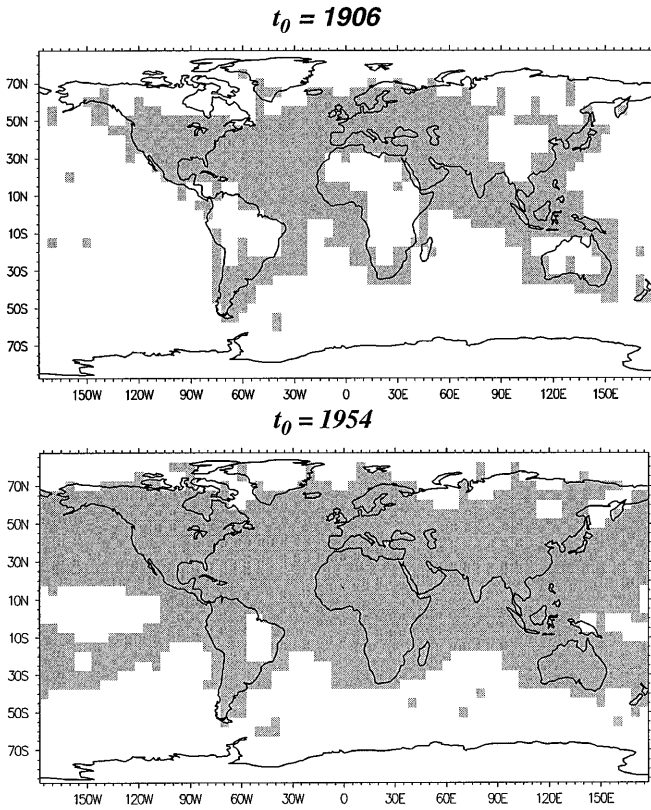


Fig. A2. Spatial coverage of annual mean near-surface temperature data for two different reference periods ($t_0=1906$ and $t_0=1954$)

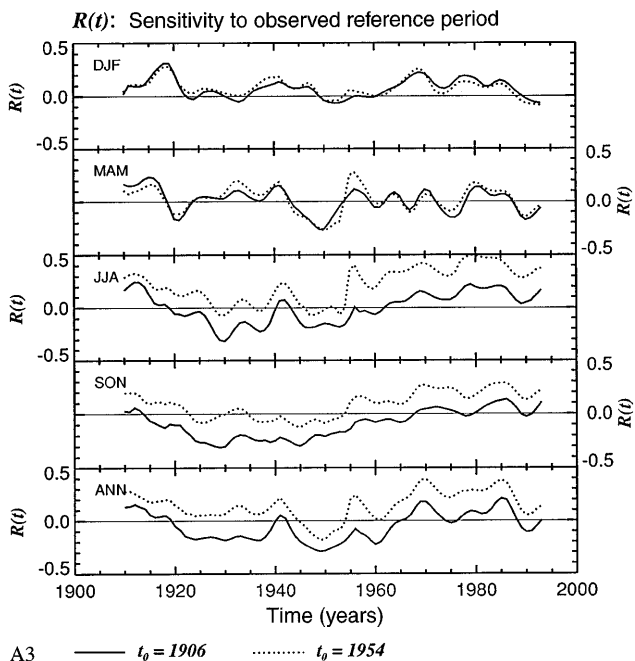


Fig. A3. Centered pattern correlation, $R(t)$, between observed seasonally and annually averaged near-surface temperature changes and changes in the TP response experiment with combined sulfate aerosol/ CO_2 forcing. Results are for two different choices of reference year: $t_0=1906$ and $t_0=1954$.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

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 Figs. A1 and A2. Since the maximum spatial coverage of observed data is determined by the coverage over the reference period (this 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 coverage of the reference state $\overline{D}_1(x)$ in order to calculate changes relative to $\overline{D}_1(x)$), changes in coverage with time will mean that the data 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 Figs. A3 and A4

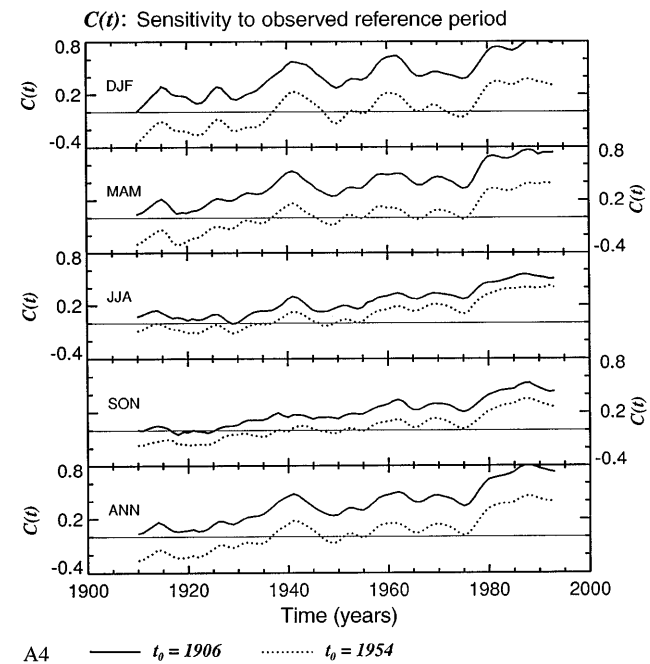


Fig. A4. As for Fig. A3, but for the uncentered pattern correlation $C(t)$, without subtraction of model and observed spatial means

for $R(t)$ and $C(t)$, respectively. We consider the extreme cases $t_0=1906$ and $t_0=1954$. Results are for the near-surface temperature signals in the TP experiment with combined CO_2 /sulfate aerosol forcing. For both statistics, it is clear that their *changes* in 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 Figs. 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).

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. (1994, 1995b) and Hegerl et al. (1994). HAMCTL exhibits substantial variability on decadal-to-century time scales (see Fig. B1a). However, part of this variability, particularly over the first 200 years of the integration, may be a model artifact and attributable to a form of climate drift.

The variability in GFDLCTL (Fig. B1b) has been

analyzed by Delworth et al. (1993) for the thermohaline circulation and by Stouffer et al. (1994), Mehta and Delworth (1995), and Manabe and Stouffer (1995) for near-surface air temperature and SST. These investigations have attempted to validate the interannual- to interdecadal time scale variability simulated by the GFDL CGCM. The study by Stouffer et al. (1994) showed qualitative agreement between observed and model-simulated patterns of interannual near-surface temperature variability (with the exception of the tropical Pacific, where both GFDLCTL and HAMCTL underestimate the variability associated with ENSO phenomena). It also indicated that modeled and observed relationships between global-mean changes and regional temperature changes show a broadly similar structure. Mehta and Delworth (1995) found considerable similarity between simulated and observed patterns characterizing decadal time scale 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 time scales. Validation of model variability on longer time scales 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 adjustment 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 Raper 1991b), the

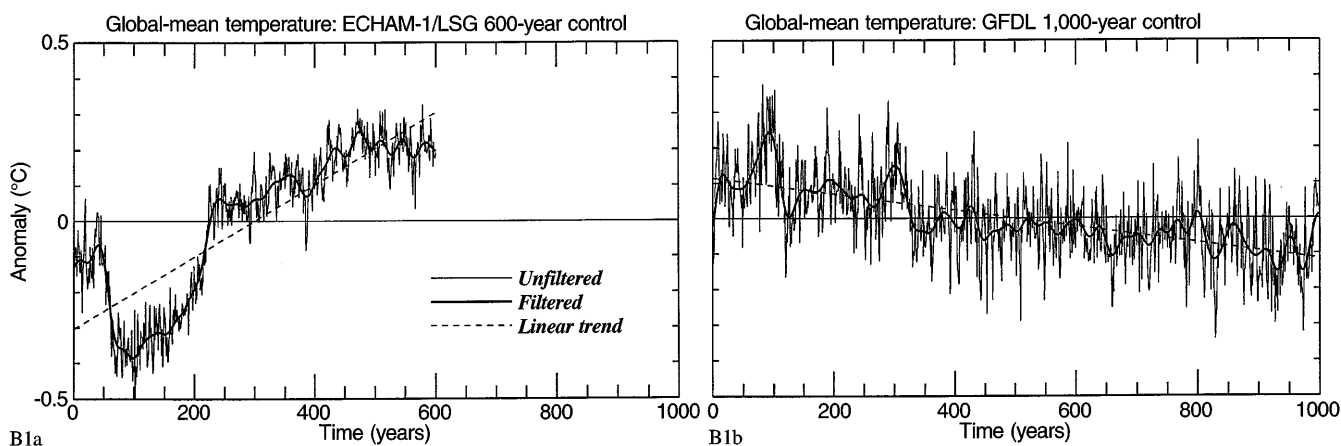


Fig. B1a, b. Time series of global-mean, annually averaged near-surface temperature changes **a** in the 600-year HAMCTL and **b** 1000-year GFDLCTL integrations. Both control runs were performed with CGCMS. Anomalies were defined relative to the mean state over the entire integration. The unfiltered and low-

pass filtered time series are shown, together with the overall 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)

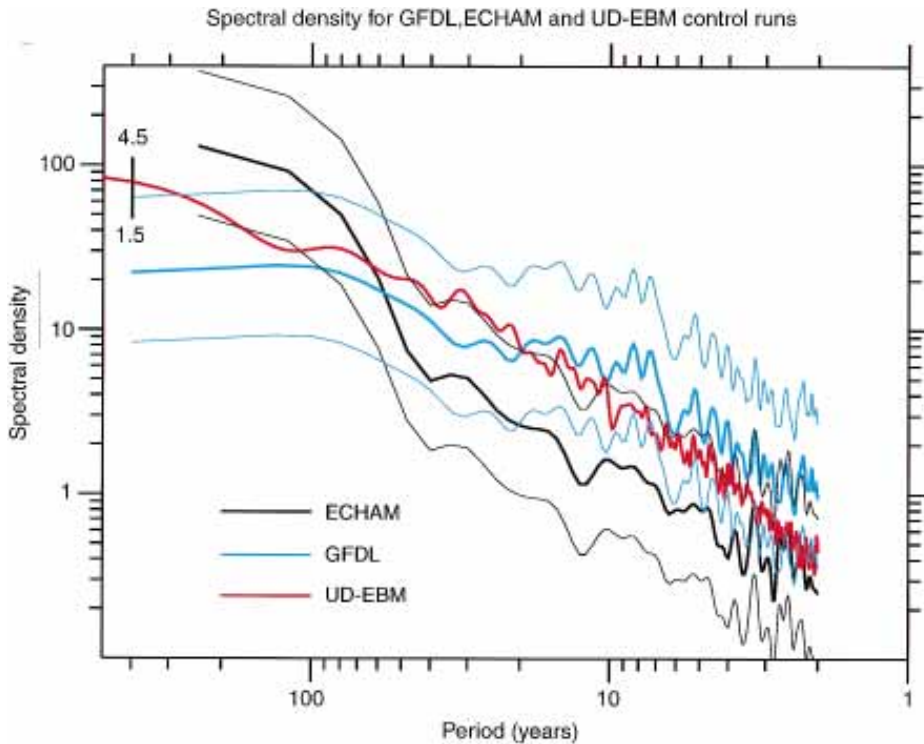


Fig. B2. Power spectra of global-mean, annually averaged near-surface temperature changes in the 600-year HAMCTL and 1000-year GFDLCTL integrations, and in a 5000-year integration performed with an upwelling-diffusion EBM (Wigley and Raper 1991b). Spectra were computed by taking the Fourier transform of the autocorrelation function, with the maximum number of lags equal to $\frac{1}{4}$ 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 year; the reverse is true for periods shorter than 60 year. 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 year. The coupled model spectra were computed after removal of overall linear trends in the data

standard deviation of the residuals is 0.10°C , a result that is robust to uncertainties in the assumed anthropogenic 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 time scales of roughly 2–60 years, and less power than HAMCTL for periods greater than ≈ 60 years (see Fig. B2). These results are relatively insensitive to the choice of window width for smoothing. 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. Removal of the overall linear trend reduces, but does not remove the effect of this initial drift on the estimated spectrum. 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 5000-year integration of an upwelling-diffusion energy-balance model (EBM; Wigley and Raper 1990; 1991b) with

random forcing and a climate sensitivity, $\Delta T_{2\times\text{CO}_2}$, of 3.0°C for a doubling of atmospheric CO_2 . As noted by Hasselmann (1976) and Wigley and Raper (1990), the low frequency spectrum depends on $\Delta T_{2\times\text{CO}_2}$. 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. At 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 on time scales less than 10 years). Note that in the lower frequency range, HAMCTL has higher power than GFDLCTL despite having a lower value of $\Delta T_{2\times\text{CO}_2}$, a result opposite to that expected on theoretical grounds, as the EBM result on the left-hand side of Fig. 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 EBM also shows a flattening of the spectrum, but at frequencies substantially

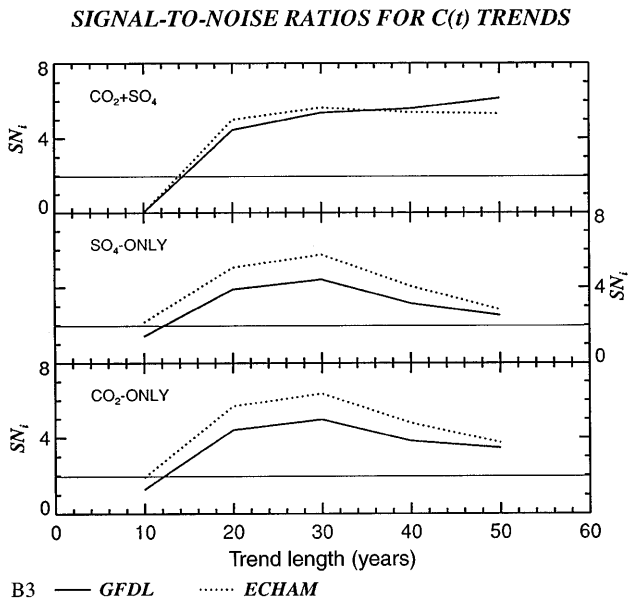
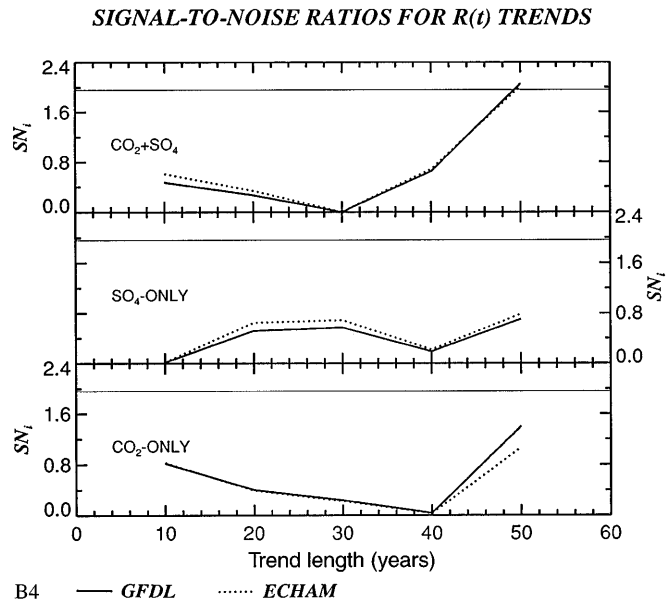


Fig. B3. Signal-to-noise ratios, SN_i , for observed trends in the un-centered pattern correlation statistic $C(t)$. The signals, $\beta_S(i)$, are the observed trends in $C(t)$ for the last 10- to 50 year. The noise, $\beta_N(i)$, is the standard error of the sampling distribution of $\beta_N(i)$ (see Sect. 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 CO_2 -only experiments, SN_i is consistently lower for the GFDLCTL noise estimates. The GFDLCTL versus HAMCTL differences in SN_i can be interpreted in terms

lower than in the GFDL model simulation (and beyond the lowest frequency shown in Figure B2). The ubiquity of substantial power at century time scales 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-time scale variability associated with internal oscillations of the coupled atmosphere-ocean system (see Barnett et al. 1995). A further possibility is that the paleodata may contain spurious low-frequency variability.

Since $C(t)$ 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 5–7 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 5–7). We therefore computed signal-to-noise ratios for $C(t)$ trends of different length. As in Santer et al. (1995b), we define the signal-to-noise ratio SN_i for trend length $i \times 10$ as

$$SN_i = \beta_S(i) / \beta_N(i) \quad (10)$$



of the coupled-model differences in the spectral density of global-mean, annually averaged temperature on time scales of 10- to 50 years (see Fig. B2). The *thin solid line* gives the 5% significance threshold for an assumed Gaussian distribution of trends

Fig. B4. As for Fig. 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 time scales

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

$$\beta_N(i) = \left[(m_i - 1)^{-1} \sum_{c=1}^{m_i} \beta_N(i, c)^2 \right]^{1/2} \quad (11)$$

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 CO_2 -only experiments, the higher GFDLCTL variance on 10- to 50-year time scales (Fig. B2) yields consistently lower signal-to-noise ratios than for HAMCTL. For trends ≥ 20 years, SN_i is much larger than the 5% significance threshold (for an assumed Gaussian distribution), so that GFDLCTL versus HAMCTL variance differences do not influence decisions on the significance of observed $C(t)$ trends. As discussed previously, in the specific case of the SC experiment $C(t)$ provides a mix of pattern- and global-mean change information. In this instance, therefore, differences in SN_i cannot be interpreted solely in terms of the different behavior of GFDLCTL and HAMCTL global-mean temperature spectra in Fig. B2.

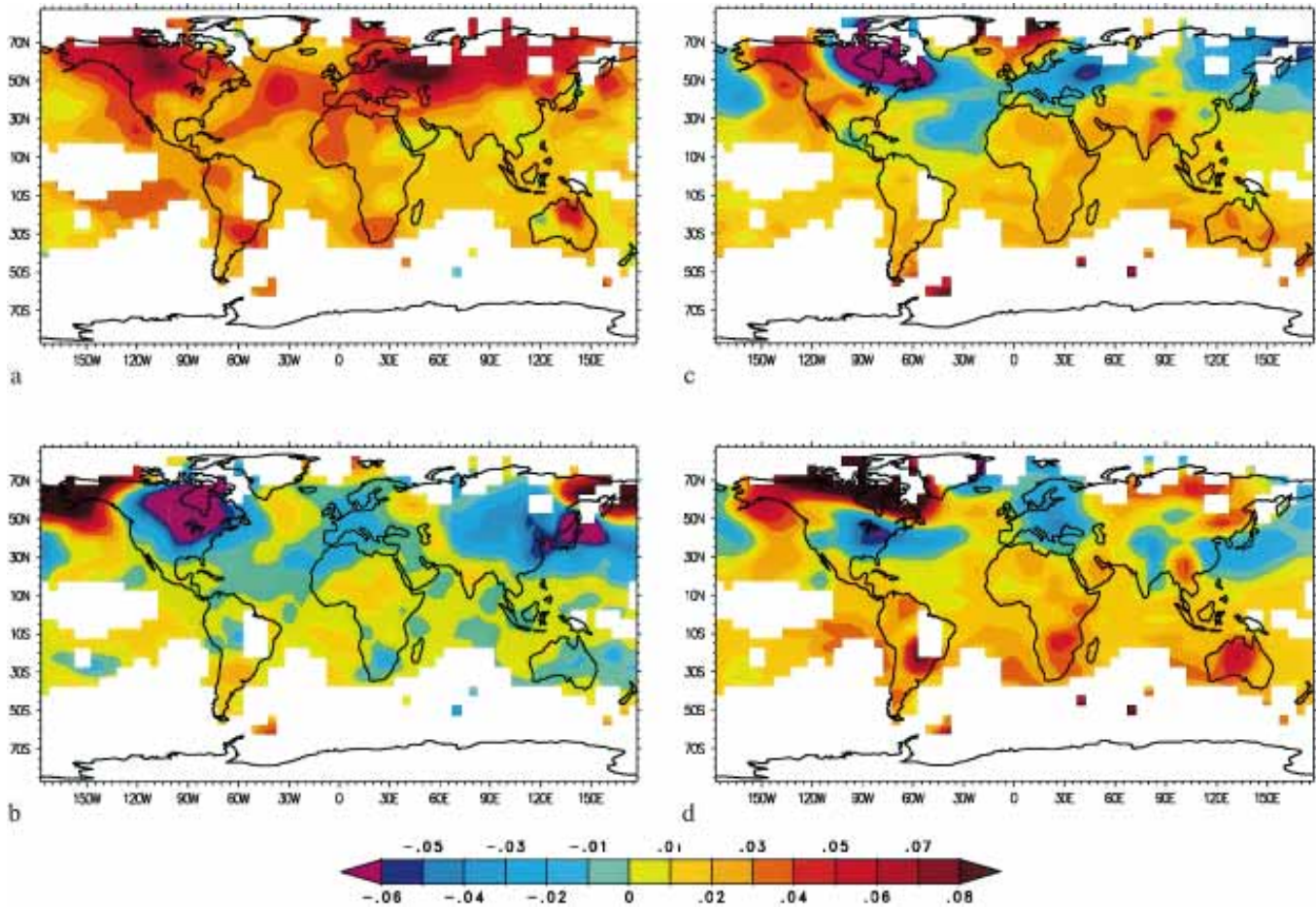


Fig. B5a–d. Empirical orthogonal functions (EOFs) 1 and 2 of the GFDLCTL and HAMCTL annually averaged near-surface temperature anomaly data. EOFs were computed using data from the first 400 year of GFDLCTL and the last 400 year of HAMCTL (in order to exclude the large negative temperature excursion during the first 200 year of HAMCTL; see Fig. B1a). Anomalies were computed relative to the 400-year means of each control run, and were filtered as described in Sect. 4. To facilitate comparison with the observed $R(t)$ results, model data were in-

terpolated to the observed grid, and grid-points outside the observed data mask for $t_0=1954$ (see Fig. 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

Signal-to-noise ratios for 10- to 50-year linear trends in $R(t)$ are surprisingly similar, at least for JJA (Fig. B4). Note that the p -values for $R(t)$ trends (Tables 5–7) are also similar for GFDLCTL and HAMCTL noise estimates. This suggests that the patterns of near-surface temperature variability on 10- to 50-year time scales must show some degree of correspondence in GFDLCTL and HAMCTL. To compare the dominant patterns of variability, we computed empirical orthogonal functions (EOFs) of the HAMCTL and GFDLCTL annually averaged near-surface temperature data (using the first 400 years of GFDLCTL and the last 400 years of HAMCTL). Model anomaly data were filtered as described in Sect. 4, and grid-points outside the observed data mask for $t_0=1954$ were excluded (see Fig. 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 control integrations. 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 GFDLCTL modes.

A 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 GFDLCTL than for HAMCTL. Further work is required to identify common patterns of variability in the

two integrations and their characteristic time scales, and to determine why the GFDLCTL and HAMCTL 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* 370:24–35
- Anderson JL, van den Dool HM (1994) Skill and return of skill in dynamic extended-range forecasts. *Mon Weather 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
- Barnett TP, Santer BD, Jones PD, Bradley RS, Briffa KR (1995) Estimates of low-frequency natural variability in near-surface air temperature. The Holocene (accepted)
- Benkovitz C (1982) Compilation of an inventory of anthropogenic emissions in the United States and Canada. *Atmos Environ* 16:1551–1563
- Bloomfield P, Nychka D (1992) Climate spectra and detecting climate change. *Clim Change* 21:275–288
- Charlson RJ, Wigley TML (1994) Sulfate aerosol and climate change. *Sci Am* 270(2):48–57
- 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
- Crowley TJ, North GR (1991) *Paleoclimatology*. Oxford University Press, New York, USA
- 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 Clim* 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
- Enting IG, Heimann H, Wigley TML (1994) Future emissions and concentrations of carbon dioxide: key ocean/atmosphere/land analyses. CSIRO Division of Atmospheric Res Techn Pap 31
- 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, pp 135–170
- 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, pp 97–134
- 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
- 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
- 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, Ruedy R, Sato M, Wilson H (1993) How sensitive is the world's climate? *Nat Geog Res Explor* 9:142–158
- Hasselmann K (1976) Stochastic climate models. Part I: theory. *Tellus* 28:473–485
- Hasselmann K (1979) On the signal-to-noise problem in atmospheric response studies. In: Shaw DB (ed) *Meteorology of tropical oceans*. Roy Meteorol Soc London, pp 251–259
- Hasselmann K (1993) Optimal fingerprints for the detection of time dependent climate change. *J Clim* 6:1957–1971
- Hegerl GC, Storch H, Hasselmann K, Santer BD, Cubasch U, Jones PD (1994) Detecting anthropogenic climate change with an optimal fingerprint method. Max-Planck Institut für Meteorologie Report 142, Hamburg, Germany
- 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, USA
- 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) Marine and land temperature data sets: a comparison and a look at recent trends. In: Schlesinger ME (ed) *Greenhouse-gas-induced climatic change: a critical appraisal of simulations and observations*. Elsevier, Amsterdam, pp 153–172
- Karl TR, Heim RR Jr, Quayle RG (1991) The greenhouse effect in central North America: if not now, when? *Science* 251:1058–1061
- Karl TR, Knight RW, Kukla G, Gavin J (1995) 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, Chichester, UK, pp 363–382
- Kelly PM, Jones PD, Sear CB, Cherry BSG, Tavakol RK (1982) Variations in surface air temperatures, Part 2: Arctic regions, 1881–1980. *Mon Weather 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 Am Meteorol 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 (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 (1995) Decadal variability of the tropical Atlantic Ocean surface temperature in shipboard measurements and in a global ocean-atmosphere model. *J Clim* 8:172–190
- Mitchell JFB, Davis RA, Ingram WJ, Senior CA (1995a) On surface temperatures greenhouse gases and aerosols: models and observations. *J Clim* (in press)
- Mitchell JFB, Johns TC, Gregory JM, Tett SFB (1995b) Transient climate response to increasing sulphate aerosols and greenhouse gases. *Nature* 376:501–504
- 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, 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 Meteorol 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 (1995) Climatic response to anthropogenic sulfate forcing simulated with a general circulation model. In: Charlson R, Heintzenberg J (eds) *Aerosol forcing of climate*. John Wiley, Chichester, UK, pp 349–362
- Santer BD, Wigley TML, Jones PD (1993) Correlation methods in fingerprint detection studies. *Clim Dyn* 8:265–276
- Santer BD, Brüggemann W, Cubasch U, Hasselmann K, Höck H, Maier-Reimer E, Mikolajewicz U (1994) Signal-to-noise analysis of time-dependent greenhouse warming experiments. Part 1: pattern analysis. *Clim Dyn* 9:267–285
- Santer BD, Taylor KE, Wigley TML, Penner JE, Jones PD, Cubasch U (1995a) Towards the detection and attribution of an anthropogenic effect on climate. PCMDI Report 21, Lawrence Livermore National Laboratory, Livermore, California
- Santer BD, Mikolajewicz U, Brüggemann W, Cubasch U, Hasselmann K, Höck H, Maier-Reimer E, Wigley TML (1995b) Ocean variability and its influence on the detectability of greenhouse warming signals. *J Geophys Res* 100:10693–10725
- Schlesinger ME, Mitchell JFB (1987) Climate model simulations of the equilibrium climatic response to increased carbon dioxide. *Rev Geophys* 25:760–798
- 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, Bryan K (1989) Interhemispheric asymmetry in climate response to a gradual increase of atmospheric CO₂. *Nature* 342:660–662
- Stouffer RJ, Manabe S, Vinnikov K Ya (1994) Model assessment of the role of natural variability in recent global warming. *Nature* 367:634–636
- Taylor KE, Ghan SJ (1992) An analysis of cloud liquid water feedback and global climate sensitivity in a general circulation model. *J Clim* 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 (1989) Possible climate change due to SO₂-derived cloud condensation nuclei. *Nature* 339:365–367
- Wigley TML (1991) Could reducing fossil-fuel emissions cause global warming? *Nature* 349:503–506
- 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, pp 471–482
- Wigley TML, Raper SCB (1991b) Detection of the enhanced greenhouse effect on climate. In: Jäger J, Ferguson HL (eds) *Climate change: science, impacts and policy*. Cambridge University Press, Cambridge, UK, pp 231–242
- 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, Raper SCB (1989) Statistical significance of global warming. *Proc 13th Ann Clim Diag Workshop*, pp A1–A8
- Wigley TML, Jones PD, Kelly PM, Hulme M (1992) Recent global temperature changes: ozone and aerosol influences. *Proc 16th Ann Clim Diag Workshop*, pp 194–202
- Woodward WA, Gray HL (1993) Global warming and the problem of testing for trend in time series data. *J Clim* 46:953–962