Climate Dynamics (2000) 16:737–754 © Springer-Verlag 2000

G. C. Hegerl · P. A. Stott · M. R. Allen
J. F. B. Mitchell · S. F. B. Tett · U. Cubasch

Optimal detection and attribution of climate change: sensitivity of results to climate model differences

Received: 30 November 1998 / Accepted: 13 December 1999

Abstract Fingerprint techniques for the detection of anthropogenic climate change aim to distinguish the climate response to anthropogenic forcing from responses to other external influences and from internal climate variability. All these responses and the characteristics of internal variability are typically estimated from climate model data. We evaluate the sensitivity of detection and attribution results to the use of response and variability estimates from two different coupled ocean atmosphere general circulation models (HadCM2, developed at the Hadley Centre, and ECHAM3/LSG from the MPI für Meteorologie and Deutsches Klimarechenzentrum). The models differ in their response to greenhouse gas and direct sulfate aerosol forcing and also in the structure of their internal variability. This leads to differences in the estimated amplitude and the significance level of anthropogenic signals in observed 50-year summer (June, July, August) surface temperature trends. While the detection of anthropogenic influence on climate is robust to intermodel differences, our ability to discriminate between the greenhouse gas and the sulfate aerosol signals is not. An analysis of the recent warming, and the warming that occurred in the first half of the twentieth century, suggests that simulations forced with combined changes in natural (solar and volcanic) and anthropogenic (greenhouse gas and sulfate aerosol) forcings agree best with the observations.

G. C. Hegerl (✉)
Texas Center for Climate Studies, Department of Oceanography, Texas A&M University, College Station, USA
E-mail: hegerl@ocean.tamu.edu

P. A. Stott · J. F. B. Mitchell · S. F. B. Tett
Hadley Centre for Climate Prediction and Research, Bracknell, UK

M. R. Allen
Rutherford Appleton Laboratory, Oxford, UK

U. Cubasch
Deutsches Klimarechenzentrum, Hamburg, Germany

1 Introduction

Model simulations predict changes in climate in response to anthropogenically induced changes in the composition of the atmosphere. In order to have more confidence in model simulations of future climate, it is useful to detect and quantify the model simulated anthropogenic climate change in observations. As pointed out by Santer et al. (1996a) and Hasselmann (1997), detection and attribution of anthropogenic climate change requires that the observations are:

1. Inconsistent with internal climate variability only. In the following, we refer to variability due to the internal dynamics of the atmosphere-ocean system as "climate noise", to distinguish it from "natural climate variability", which encompasses climate noise and additionally the climate response to naturally occurring climate forcings, for example volcanism;
2. Consistent with an assumed combination of external forcing mechanisms which includes anthropogenic forcing, (for example to the increase in greenhouse gases and sulfate aerosols) and;
3. Inconsistent with other conceivable explanations of externally forced climate change (for example, combinations of natural climate forcing alone).

In this context, "inconsistent" implies that a disagreement between model and observed data is larger than the uncertainties in the observations and in model-based estimates of the climate noise and expected climate response.

Many studies have shown that the most recent observations of changes in surface and free-atmosphere temperatures cannot be explained by (model-estimated) natural climate variability alone. The observations agree best with model simulations that account for changes in greenhouse gases and sulfate aerosols, and possibly also stratospheric ozone (for example, Santer et al. 1995, 1996a, b; Mitchell et al. 1995a, b; Tett et al. 1996; Hegerl et al. 1996, 1997; North and Stevens 1998; Wigley et al.

1998; Tett et al. 1999; Stott et al. 2000). All these approaches rely on data from climate models to estimate both the space-time pattern of the climate response to external forcings ("climate change signals") and the magnitude and patterns of internal climate variability. Only some authors check for the validity of the estimate for climate variability. However, a more rigorous detection and attribution of climate change would have to take model errors, errors in the forcing and errors in the observations into account.

We investigate the impact of uncertainties in model-based estimates of anthropogenic climate-change signals and internal climate variability on optimal detection results. We do this by applying data from two different models in an optimal detection method. We identify the major model differences in these estimates, and examine their effect on detection and attribution results.

Optimal detection methods are powerful tools for detecting climate change and attributing it to a combination of external forcings (Hasselmann 1979, 1993, 1997; Bell 1986; North et al. 1995; Allen and Tett 1999). However, optimal detection methods are also especially sensitive to model error. The reason is that not only the climate change pattern, but also the covariance of climate noise (required for computing optimal fingerprints) is derived from climate model data. The climate change signal together with the noise covariance yields an optimal fingerprint which helps to focus mostly on those features of the climate change signal which are most distinct from noise. Therefore, optimal fingerprints produced from different signal patterns and different noise will yield at least quantitatively different results.

We use two sets of optimal fingerprints derived from two different global coupled ocean-atmosphere general circulation models (OAGCMs): The "Had" fingerprints, where the signal patterns, and the estimate of the noise-covariance were derived from simulations with the Hadley Centre's HadCM2, and the "MPI" fingerprints, where the signal patterns and noise covariance were derived from simulations with the ECHAM3/LSG coupled model. We use both sets of fingerprints to first estimate the observed amplitude of the combined effect of greenhouse gas and sulfate aerosol forcing, and then apply a two-signal analysis using a greenhouse gas only and a sulfate only signal pattern. Independent data from the ECHAM3/LSG and HadCM2 control simulations are used in both cases to assess signal amplitudes occurring due to climate noise only.

The individual results from both models have been used previously for optimal detection studies. Results from various versions of the MPI coupled models have been used to detect climate change in surface temperature, demonstrating that recent observed 30-year trends are significantly stronger than trends occurring due to internal climate variability (Hegerl et al. 1996; Hegerl et al. 1997). This conclusion did not depend on whether the model simulations involved changes in greenhouse gas alone or also changes in sulfate aerosols. It was also independent of which of three coupled models was used to estimate the covariance structure of climate noise. The stability of this result is mainly due to two factors: the strength of the observed global mean warming trend over the recent decades, which dominates smaller scale pattern information, and the use of statistics that preserve the spatial means of the fields being compared (Santer et al. 1995; Hegerl et al. 1996; Stott and Tett 1998).

Further work with MPI model simulations has shown that the recent observed 50-year temperature trends in summer are in closest agreement with simulations incorporating combined greenhouse gas and sulfate aerosol forcing. The observations are inconsistent with climate noise and the simulated response to greenhouse gas forcing only. Additionally, the observations are marginally inconsistent with climate change due to solar forcing alone (Hegerl et al. 1997; henceforth H97).

Tett et al. (1999, in the following, T99) and Stott et al. (2000; in the following, S00) have used the HadCM2 simulations to estimate the amplitude of the model's time-space climate-change pattern in observed near-surface temperature records. They demonstrated that the observations contain a significant greenhouse gas-plus-sulfate aerosol signal, and also found some evidence for a solar forced component in the first half of the century. There are a number of technical differences between the approach used in T99 and S00 and that employed here and in H97.

The main difference is that while T99 and S00 use the full time-space evolution of decadal mean data over 50-year periods in their optimization strategy, the present work and H97 rely on patterns of 50-year summertime temperature trends (JJA). In the latter case, the optimization is in space only and does not explicitly use information on the temporal evolution of the predicted climate-change signal. This has disadvantages in distinguishing spatially correlated signal patterns with different time history (e.g., signals for sulfate and greenhouse gas, or solar and greenhouse gas forcing). However, for intercomparisons, this has the advantage that model differences in temporal behavior need not be considered in attempting to understand model differences in optimal fingerprints.

The comparison between the results obtained with the Hadley Centre (Had) and MPI fingerprints (MPI) gives an indication of the sensitivity of earlier detection and attribution results to model signal uncertainties. It also helps to highlight current uncertainties in model estimates of the response to anthropogenic forcing.

The work is organized as follows: the optimal detection and attribution method is briefly outlined in Sect. 2. Section 3 discusses the observations, the climate change simulations and the estimates of climate noise. Differences in climate change patterns and climate noise are assessed in Sect. 4, while Sect. 5 discusses the results of applying the detection and attribution method using fingerprints from different models. Conclusions are given in Sect. 6.

2 Optimal detection and attribution method

The theory of detection and attribution using optimal fingerprints is well established (we use notation as in Hasselmann 1979, 1993, 1997; Allen and Tett 1999). Here, only those aspects of the method are briefly outlined which are required to understand this work. The method is based on estimating the amplitudes of one or more patterns, $g_v, v = 1, 2, \ldots, p$, of response to external forcing(s) in the observed pattern of climate change from the observed climate state $\boldsymbol{\Psi} = (\boldsymbol{\Psi}_i)_{i=1\ldots n}$:

$$\boldsymbol{\Psi} = \sum_{v=1}^{p} a_v g_v + \tilde{\boldsymbol{\Psi}}_r \ . \tag{1}$$

a_v denotes the amplitude of each signal guess pattern g_v in the observations. If all forcings are taken into account, $\tilde{\boldsymbol{\Psi}}_r$ is a noise residual. The best (least noise influenced) estimate of the amplitudes a_v can be obtained by a multivariate regression using the inverse noise covariance in the scalar product (Hasselmann 1979, 1997, H97; Allen and Tett 1999; see also North et al. 1995). We focus on the case that climate change signals are orthogonal to each other (relative to the same scalar product),

$$g_v^T C^{-1} \cdot g_\mu = \delta_{v\mu}, \quad v, \mu = 1, \ldots, p, \quad \text{with} \quad (C_{ij})_{i,j=1,\ldots,n} = \langle \tilde{\Psi}_i \tilde{\Psi}_k \rangle \ . \tag{2}$$

where $\tilde{\Psi}_i$ is a realization of climate noise and "$< \cdot, \cdot >$" denotes the expectation.

Then the amplitudes a_v of signals are computed by

$$a_\mu = g_\mu^T C^{-1} \boldsymbol{\Psi} \ . \tag{3}$$

$f_{-\mu} = C^{-1} g_\mu$ is referred to as "optimal fingerprint". The use of greenhouse gas and orthogonalized sulfate fingerprints represents a stepwise regression. This is appropriate when the second signal represents a perturbation of the first. Allen and Tett (1999) and T99 use a multiregression procedure which treats all signals identically. Tests for consistency between simulations and observations are identical in the two approaches (Hegerl and Allen, in prep.).

The process of detection and attribution of anthropogenic climate change consists of comparing the estimated vector of amplitude estimates of the climate change signals from the observations $\widehat{\boldsymbol{a}}_{obs} = (\widehat{\boldsymbol{a}}_{obs})_{v=1:p}$ with that from model simulations run with different forcing histories $\widehat{\boldsymbol{a}}_{M1}, \ldots, \widehat{\boldsymbol{a}}_{Mn}$. Both the estimated amplitudes of climate change patterns in observed and model data deviate from the 'true' underlying amplitudes $\boldsymbol{a}_{obs}$ and $\boldsymbol{a}_{M1}, \ldots, \boldsymbol{a}_{Mn}$ due to the following uncertainties:

1. O1: climate noise influence on observed pattern amplitude estimates $\widehat{\boldsymbol{a}}_{obs}$. Even if optimal detection methods are applied, noise cannot be completely separated from the signal.
2. O2: observational sampling error, which affects $\widehat{\boldsymbol{a}}_{obs}$. Sampling error occurs since an observed gridbox value is the average of a limited number of stations, and since trend values at a gridbox are computed even if individual months and single seasons are missing (see Hegerl et al. 2000).
3. O3: instrumental error affecting $\widehat{\boldsymbol{a}}_{obs}$, e.g., measurement errors and uncertainties in corrections applied for inhomogeneities.
4. M1: sampling noise in the model estimated climate change signal, which depends on the number of simulations which have been averaged and the statistics of the model's noise. M1 influences the model amplitude $\widehat{\boldsymbol{a}}_M$, and, through noise in the fingerprint also $\widehat{\boldsymbol{a}}_{obs}$.
5. M2: systematic errors in the model signal pattern and its amplitude due to model errors affect both $\widehat{\boldsymbol{a}}_M$ and, through the fingerprint, also $\widehat{\boldsymbol{a}}_{obs}$.
6. M3: uncertainties and errors in the forcing of a model simulation also affect both $\widehat{\boldsymbol{a}}_M$ and $\widehat{\boldsymbol{a}}_{obs}$.

Climate change is "detected" if the null hypothesis $\boldsymbol{a}_{obs} = 0$ is rejected given $\widehat{\boldsymbol{a}}_{obs}$ and all uncertainties that affect $\widehat{\boldsymbol{a}}_{obs}$. Note that for that detection test, M1 to M3 yield an underestimate of signals and hence conservative results (see Hegerl and North 1997; Allen and Tett 1999). The attribution test ideally assesses if the differences between the observed amplitude estimates $\widehat{\boldsymbol{a}}_{obs}$ and that in each of the model simulations $\widehat{\boldsymbol{a}}_M$ can be explained by all combined uncertainties O1–O3 and M1–M3 (Hasselmann 1997). For that purpose, it is assessed if the difference $\widehat{\boldsymbol{a}}_{obs} - \widehat{\boldsymbol{a}}_{Mi}$ is consistent with the uncertainty in the amplitude estimates from the model simulation and the observations. The null hypothesis is tested that both amplitudes are identical, and the difference is due to the above listed uncertainties only. If this hypothesis is rejected at a given significance level for a model simulation's amplitude estimate, the model simulation, and along with it, the associated forcing mechanism is rejected as a plausible cause for the observed climate change signal at the specified risk (significance) level. Otherwise, no significant disagreement has been found between the observations and the model simulation and the forcing mechanism is considered a plausible cause for climate change.

H97, T99 and S00 incorporate O1 in the attribution exercise. Jones et al. (1997) have estimated observational sampling error in O2, which Hegerl et al. (2000) have applied in an optimal detection study. Results indicate that the effect of sampling uncertainty is small if areas not covered by observations are excluded from the detection and attribution exercise (Hegerl et al. 2000). For conservativeness, we incorporated the effect of instrumental sampling error in the present study. North and Stevens, (1998) avoid part of the sampling error by reducing their analysis to a number of continuously reporting gridboxes. Instrumental error O3 cannot be fully addressed at present, but is estimated as small, particularly for recent trends (Jones et al. 1999 see also Hegerl et al. 2000).

H97, T99 and S00 also incorporated the effect of M1 on $\widehat{\boldsymbol{a}}_M$. For M2 and M3 only some sensitivity studies have been performed: Santer et al. (1996b) used signals from different models in pattern correlation studies. Barnett et al. (1998) have studied model uncertainty in atmospheric GCMs by comparing the differences between observations and simulations with several atmospheric GCMs forced with observed sea-surface temperature. They found temporal and spatial coherence in the differences, which however, did not project strongly onto anthropogenic signals on multidecadal time scales (note that errors which differ from the signal shape could potentially also have a severe effect on distinguishing between several, similar signals). Barnett et al. (1998) could not separate true model error from atmospheric model noise, and hence determine the covariance structure of the model error. Tett et al. (1996) and Santer et al. (1996b) included sensitivity tests regarding the effect of stratospheric ozone forcing. A systematic treatment of pattern errors due to M2 and M3 is not possible at present.

This study systematically investigates the effect of the uncertainties in the model response to anthropogenic forcing (M2). This is done by constructing different sets of optimal fingerprints using simulations from two different models run with similar anthropogenic forcing. Amplitude estimates from these simulations are compared to the observations using both sets of optimal fingerprints. Both signal and noise uncertainties influence the optimal detection results (Eq. 3), as can be demonstrated if we assume that the signal $g_\mu = (g_{\mu 1}, \ldots, g_{\mu n})$ and the observations $\boldsymbol{\Psi} = (\psi_1, \ldots, \psi_n)$ have been transformed into the space of the eigenvectors of the noise covariance matrix C (Eq. 2), making it a diagonal matrix ("EOF-coordinates"). The amplitude estimate for each signal (assuming again that signals are orthogonal) can then be written as

$$a_\mu = \sum_i \frac{g_{\mu i} \cdot \psi_i}{\lambda_i} \ , \tag{4}$$

with λ_i denoting the eigenvalues of C. Using non-orthogonal fingerprints will yield more complicated algebra, but a similar dependence. The division by the eigenvalues of climate noise ensures that signal components associated with little noise (small λ_i) are emphasized relative to those with large noise contribution. Equation 4 shows that both the signal vector and the climate noise covariance influence the detection variable and therefore the estimated amplitude of the signal in the observations.

3 Data

3.1 Observations

We use an improved dataset of near-surface temperature observations (Parker et al. 1994; Jones 1994a, b). This is an update of the dataset used previously in H97 (Jones and Briffa 1992). The data are compiled as sea surface temperature and near-surface temperature anomalies with respect to 1961–1990 on a 5° longitude by 5° latitude grid. Differences between the earlier and later version are generally small and are related to improved gridding techniques and to changes in quality control techniques and the selected reference period. This leads to small coverage differences. In order to keep results comparable to the previous results, data coverage of the previous analysis has been only modified by excluding gridpoints which are no longer covered in the updated dataset. Results proved to be relatively insensitive to slight changes in coverage.

3.2 Climate model simulations

Output from three different global coupled OAGCMs have been used. Both modeling centers now have improved models (Collins et al. 2000; Gordon et al. 2000; Roeckner et al. 1999), but at the time of writing, these had not been used to produce the long control simulations and ensembles of anomaly simulations which facilitate the accurate estimates of the model's noise and response characteristics that are needed for optimal detection techniques.

All model data have been bi-linearly interpolated to the same grid as the observations prior to all analyses shown here.

1. ECHAM3/LSG (''HAM3L''; Voss et al. 1998): this model has an updated atmospheric component (ECHAM3) and is otherwise similar to the ECHAM1/LSG model (Cubasch et al. 1992), which features a spectral resolution of T21 (approximately 5.6° by 5.6° degrees). We use data from climate change simulations and a long control simulation (more than 1500 years). Some uncertainty is introduced by a drift in the HAM3L control simulation. The drift appears to originate in the ice edge area of the Southern Hemisphere and causes a global mean warming of 0.5 °C in the first 300 years and a slight drift afterwards. Since the drift is strongest in the first 160 years, these have been excluded from the analysis to avoid that the drift influences the noise estimate from HAM3L. For the purpose of comparison with previous results (H97) we also refer to the older ECHAM1/LSG control simulation (''HAM1L'', von Storch et al. 1997).
2. HadCM2 is the second cycle of the Hadley Centre climate model (Mitchell et al. 1995a; Johns et al. 1997; Tett et al. 1996). The HadCM2 model has a finer resolution of 2.5° latitude × 3.75° longitude than the Hamburg models. The control simulation is over 1700 years long (Tett et al. 1997). Four-member ensembles of four climate change simulations have been used in the present study.

The following anthropogenic climate change simulations have been performed with one or both models:

Climate change due to greenhouse gases

One simulation with HAM3L and an ensemble of four with HadCM2 have been forced with observed ''equivalent'' CO_2 increases from the previous century to the present, and slightly different scenarios of greenhouse gas forcing thereafter (for HAM3L the IPCC 1990 scenario A, which is approximately a 1.3% increase, see Cubasch et al. 1996, and for HadCM2 a 1% increase, Mitchell and Johns 1997). In the following, the ensemble means for these integrations are denoted by ''G''. The HadCM2 results are averaged over four realizations, while the HAM3L result is based on a single integration.

Climate change due to greenhouse gases and anthropogenic sulfate aerosols

The ensemble means of two simulations with HAM3L and four with HadCM2 are referred to as ''GS'' (Mitchell and Johns 1997; Cubasch et al. 1996). The greenhouse gas forcing is the same as in G, and the direct effect of sulfate aerosols has been taken into account through a change in annual mean surface albedo in both models. The aerosol forcing data have been provided by Langner and Rodhe (1991). While ECHAM3/LSG uses decadal mean forcing fields, with a spatially varying pattern (Cubasch et al. 1996), HadCM2 uses the 1990 forcing pattern scaled to the total annual forcing (Mitchell and Johns 1997), however, the differences are small. Note that there is a large uncertainty in the sulfate aerosol forcing on climate (Penner et al. 1995; Haywood et al. 1997). The direct sulfate aerosol forcing in both models is approximately −0.6 to −0.7 W/m^2 for the present. This is larger than the IPCC best estimate of −0.4 W/m^2, but within the given uncertainty range (IPCC 1996). Note that detection of a sulfate pattern in observations should be insensitive to the pattern amplitude in the model and therefore should not be sensitive to this uncertainty. In optimal detection, the magnitude of the sulfate pattern is estimated from observations, and the orthogonal sulfate pattern is independent from the relative magnitude of sulfate to greenhouse gas forcing in the model (disregarding nonlinear effects, which are found to be small, Haywood et al. 1997). Both models neglect the indirect effect of sulfate aerosols on cloud formation, which is highly uncertain in pattern and amplitude (Boucher and Lohmann 1995; Jones et al. 1994; IPCC 1996). This effect has been incorporated in some recent simulations (Roeckner et al. 1999; Tett et al. in preparation).

We would also like to determine whether the response to natural forcings projects on anthropogenic climate change signals and therefore might, if not taken into account, lead to errors in estimates of anthropogenic signals. The following simulations have been used:

Solar forced climate change

We analyzed the ensemble mean of two simulations of HAM3L (Cubasch et al. 1997) and four of HadCM2 which were forced with estimated change in solar radiation according to Hoyt and Schatten (1993) and are referred to as ''Sol''. The HAM3L simulations end in the last year of the Hoyt and Schatten forcing data (1992), while the HadCM2 simulations have been continued using data from Willson (1997) until 1995. Note that also for this forcing there is a very large uncertainty in the correct magnitude and time evolution of solar radiation changes before the satellite period (see Lean et al. 1995).

Climate change forced by explosive volcanism

Four simulations of HadCM2 are based on the stratospheric aerosol forcing as estimated by Sato et al. (1993), whose average is referred to as ''Vol''. Again, the exact magnitude of the forcing is uncertain, although the timing of the eruptions and the explosivity is approximately known. Some uncertainty is introduced by a drift in Vol which is probably caused by the net negative forcing, while the model is tuned to be drift free in its control climate, which has no background stratospheric aerosol levels.

The detection and attribution results based on these simulations are subject to several caveats, which need to be kept in mind: first, with the exception of well-mixed greenhouse gases, there are relatively large uncertainties in radiative forcing. Second, the assessment of consistency between simulations and observations depends on estimates of the climate's internal variability. The realism of the model noise on the required 50-year time scale cannot yet be assessed with confidence; differences between noise estimates from different models (see Figs. 4, 6, Hegerl and North 1997; Stouffer et al. 2000) show that the reliability of model noise estimates is uncertain at present. Paleodata might provide some additional

information (e.g. Jones et al. 1998, Mann et al. 1998) however, they reflect both externally forced and internal climate variability. Also, uncertainties in paleo records suggest some caution in their interpretation (see Barnett et al. 1996, Jones et al. 1998).

4 Model differences

4.1 Climate change patterns

We have derived spatial patterns that represent expected climate change from the G and the GS simulations. In the present approach (as in H97) we performed an EOF analysis on summertime temperature data from the anthropogenic climate change simulations in order to separate the underlying climate change signal from climate noise. It was found that the time evolution of the first EOF describes the most important space-time features of the signal and effectively filters out much of the noise for the GS signal in a similar manner as for the G signal (Cubasch et al. 1992; Santer et al. 1994, H97), reducing the effect of error M1 on the signal and hence on $\widehat{\boldsymbol{a}}_{obs}$. The disadvantage is that the first EOF also tends to be dominated by the 21st century response, when forcing and signal are largest, causing errors in present climate change signals if there are time-space-varying components of the forcing and climate response (see Santer et al. 1996c). More discussion of using the first EOF as a signal pattern can be found in the Appendix. T99 and S00, in contrast, use present day climate signals from an ensemble of simulations in the same manner as from the observations. This should yield a more realistic representation of the present climate change signal. However, a large ensemble size is required to effectively remove superimposed climate variability. T99 note that residual noise in the signal pattern (which are derived from an average of four realizations) can lead to a low bias in the estimated amplitudes of the climate change signals (Allen and Tett 1999), which can be addressed by a total least square approach (Allen and Stott 2000, Allen et al. 2000). Using contemporary signal patterns is less feasible for HAM3L data, since the ensemble sizes are even smaller (two and one respectively). This approach has therefore not been applied here.

For greenhouse gas forcing only, the HAM3L signal pattern shows the most pronounced warming over the northern midlatitude continents and shows a rather large region of little warming in the North Atlantic (Fig. 1). The HadCM2 signal pattern shows cooling in a more localized area of the North Atlantic, and also enhanced warming over the continents, however, with less emphasis on the northern midlatitudes (Fig. 1). In HadCM2, the tropical oceans warm more strongly than in HAM3L, while the Southern Hemisphere oceans warm less. Also, the North Pacific warms less, which may be caused by a relatively strong negative cloud-feedback in that model (Senior 1999). The greenhouse gas and sulfate aerosol signal pattern in HadCM2 (Fig. 2) shows reduced warming in the NH midlatitudes as compared to the greenhouse gas only simulation for

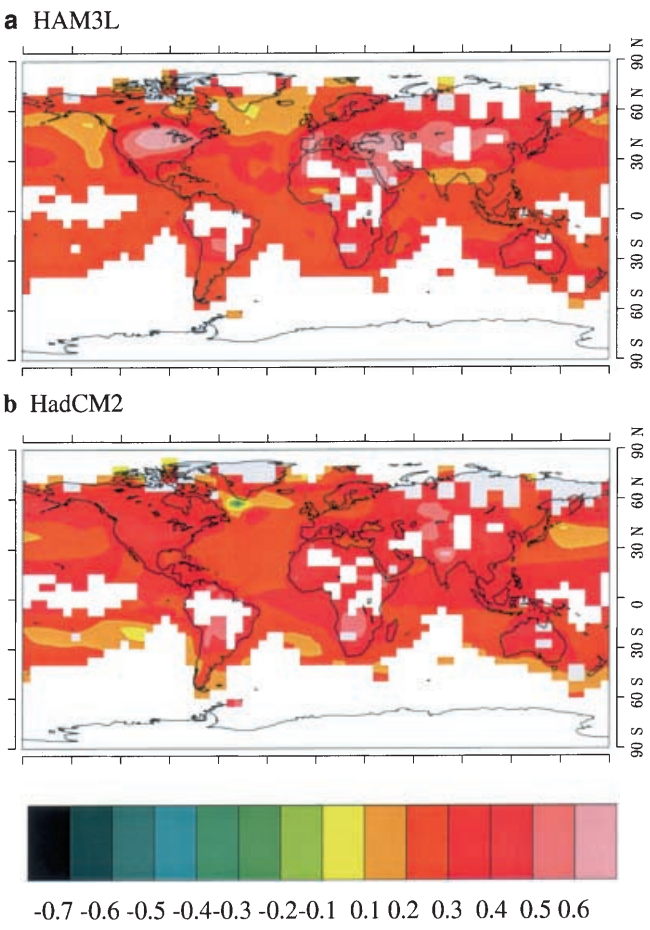

Fig. 1a, b Signal pattern of greenhouse gas only climate change (G) in **a** ECHAM3/LSG ("HAM3L") and **b** HadCM2. Only gridpoints are shown (and were used in the analysis) which are sufficiently covered by observations in the recent 50 years. Both patterns are based on an EOF analysis of each model's greenhouse gas simulation and normalized to the same total variance over the areas covered with observations. Note the pattern differences in the Southern Hemispheric ocean, the tropics, the midlatitude Pacific and Atlantic and the Northern Hemisphere midlatitude landmasses

both models. However, the differences between the G and GS spatial patterns are much larger in HAM3L than in HadCM2 as also indicated by their pattern correlation (correlation without spatial mean $r^* = 0.3$ compared to $r^* = 0.7$, Table 1). HadCM2's GS and G patterns are more similar to each other than to the respective HAM3L signal with the same forcing ($r^* \sim 0.5$). The HadCM2 signal patterns used here are very similar to those discussed in Mitchell and Johns (1997). Räisänen (1997, 1999) discusses physical differences between both models' anthropogenic climate change simulations and speculates that in the large warming in NH midlatitudes in the greenhouse gas only simulation with HAM3L is related to a drying of soil moisture in summer, which is absent under combined greenhouse gas and sulfate forcing and in the HadCM2 simulations.

We have not used fingerprints for the natural climate forcings (solar and volcanic forcing). As mentioned, we

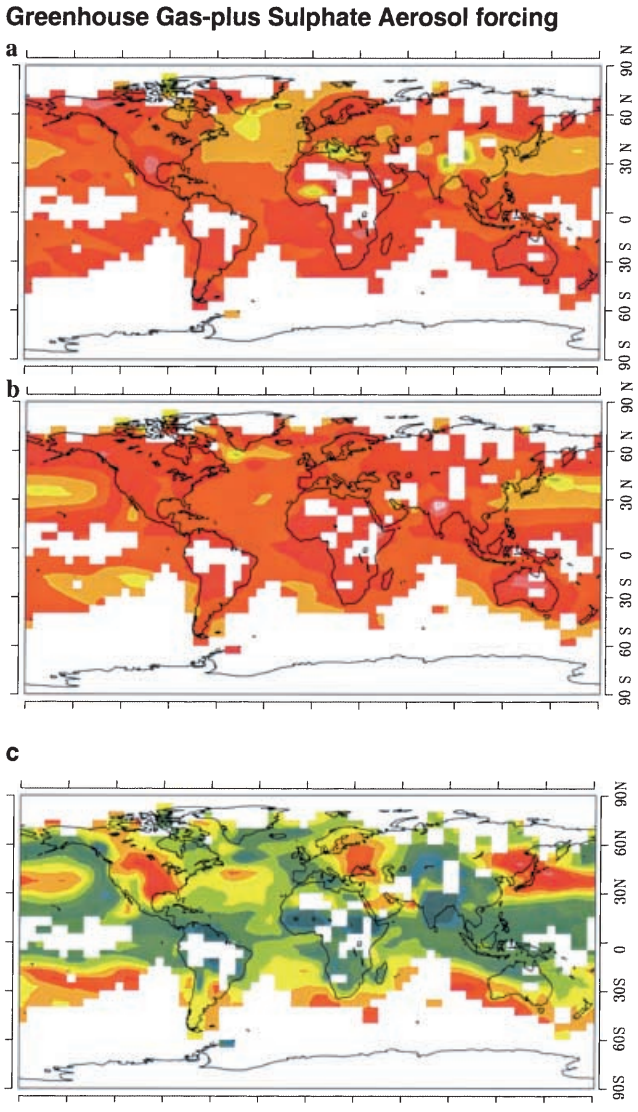

Fig. 2a–c Same as Fig. 1, but for the greenhouse gas-plus-sulfate aerosol pattern (GS). The *bottom panel* c shows the GS pattern from the HAM3L simulation, after it was orthogonalized (Eq. 2) relative to the HadCM2 pattern. This pattern focuses on the difference between the HAM3L pattern and the HadCM2 pattern (i.e., on the parts of the HAM3L pattern not represented by the HadCM2 pattern). The most prominent features are enhanced warming in the HAM3L signal pattern in the Southern Hemisphere Oceans and the North Pacific, compared to reduced warming in the tropics and large parts of the Northern Hemisphere midlatitudes

want to study the naturally forced climate signals only in terms of their influence on estimates of the anthropogenic signals. Additionally, the natural climate forcing is smaller than the anthropogenic forcing on long time scales, leading to a stronger influence of climate noise on the signal patterns and problems in deriving them. Cubasch et al. (1997) found no substantial differences in the spatial pattern of surface air temperature in response to solar and greenhouse gas forcing. We suspect that the derivation of a volcanic signal would be similarly difficult.

4.2 Internal climate variability

In all fingerprint detection and attribution approaches, the differences in the signal patterns affect estimated signal amplitudes while between-model differences in climate noise affect the estimated significance level of a detected climate change. In optimal detection methods, noise uncertainties additionally affect the estimate of signal amplitudes through the estimate of the noise covariance (see Eq. 4). Differences between annual mean surface temperature variability in two of the long control simulations applied here have been investigated in Stouffer et al. (2000). The authors found that the most general features of variability, for example, the enhanced variability over NH continents, were quite similar between the coupled climate models (note that differences between the models increased for longer time scales). To the extent that model-simulated variability reflects variability in the real climate system, using the models inverse covariance for the regression norm should therefore give more stable and efficient estimates than using an Euclidean norm (Eq. 3), where noisy features are associated with the same weight as stable features. However, the authors also found differences: HadCM2 shows a generally higher level of variability than HAM3L, which in turn was found to be more active than HAM1L (H97) and somewhat less active than the observations. Also, a common EOF analysis of 5-year averaged data showed that the variance associated with dominant patterns of variability varies between the models (Stouffer et al. 2000; see references for common EOFs therein). Such a difference would introduce different eigenvalues λ_i, and therefore differences in the amplitude estimates (Eq. 4). Here, a common EOF analysis of 50-year JJA trend patterns from the HadCM2, HAM3L and HAM1L control simulations has been performed. Prior to the analysis, all data have been transformed to the observational grid; areas not covered by observations have been excluded. Then, 650 running trend patterns (separated by 5 years) from each model have been concatenated and an EOF analysis has been performed on the concatenated and area weighted data. Within the EOF analysis, data from each model have been scaled to the same amount of total spatial variance; the scaling has been removed again prior to computation of the variance associated with each EOF pattern. This scaling avoids the most active model dominating the spatial EOF patterns, but does not affect results in a fundamental way.

Figure 3 shows the spatial pattern of the first two common EOFs. The first common EOF shows a spatial warming pattern focusing on the Northern Hemispheric midlatitudes and the North Atlantic, with cooling centered over the Aleutian Islands. The second EOF shows enhanced warming in the Pacific Ocean and the tropical continents, with slight cooling in North Pacific midlatitudes. Its pattern resembles observed ENSO-like decadal variability (Zhang et al. 1997). EOFs 3 and 4 focus on variability in NH midlatitudes, particularly over the continents (not shown). Note that the second EOF projects onto the G and GS

Table 1 Pattern correlations (upper diagonal: *r* spatial mean included, lower diagonal: *r** spatial mean subtracted; both area weighted) between greenhouse gas only G, greenhouse gas-plus sulfate aerosol, GS and orthogonal sulfate Sul climate change signals and EOFs 1 and 2 of the common EOF analysis. Correlations with EOFs 3 and 4 are small (<0.25) and not shown. Correlations larger than ± 0.5 are shown bold. „MPI" refers to patterns from HAM3L, „Had" to HadCM2 patterns

	G Had	G MPI	GS Had	GS MPI	Sul Had	Sul MPI	EOF 1	EOF 2
G Had	**1**	**0.93**	**0.93**	**0.91**	−0.06	0.08	0.29	**0.69**
G MPI	**0.53**	**1**	**0.80**	**0.90**	−0.27	0.01	0.48	**0.56**
GS Had	**0.70**	−0.03	**1**	**0.89**	0.25	0.23	0.00	**0.73**
GS MPI	0.36	0.31	**0.51**	**1**	−0.01	0.31	0.15	**0.65**
Sul Had	0.18	−0.37	**0.72**	0.35	**1**	0.20	**−0.78**	0.14
Sul MPI	−0.11	−0.28	0.24	**0.51**	0.23	**1**	−0.17	0.18
EOF 1	0.02	0.53	**−0.54**	−0.34	**−0.78**	−0.22	**1**	0
EOF 2	0.26	−0.17	0.42	0.13	0.31	0.13	−0.28	**1**

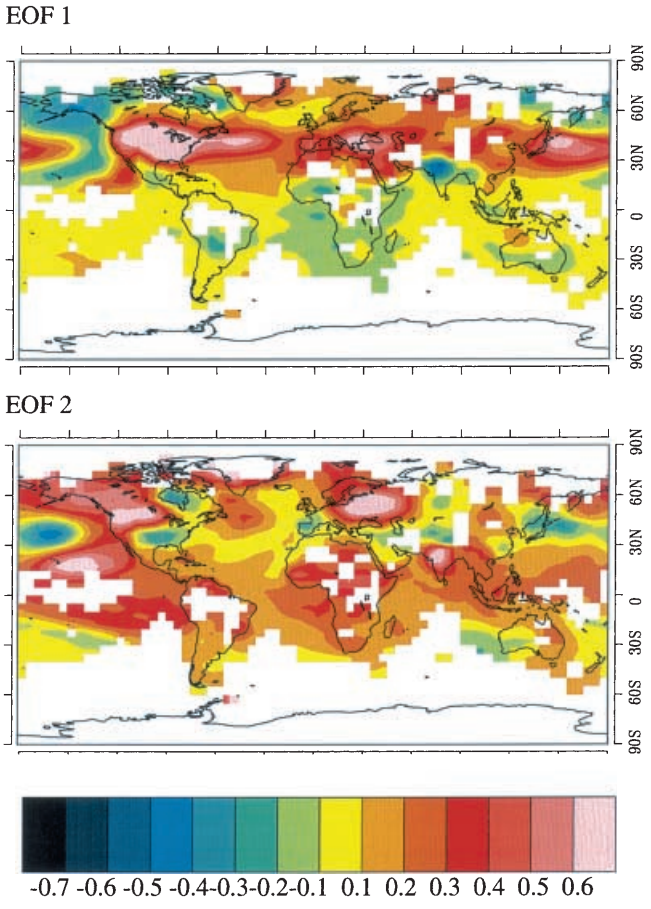

Fig. 3 Common *EOFs 1* and *2* from a concatenation of 50-year NH summer trends from the HAM3L, HadCM2 and ECHAM1/LSG control simulations. EOF 2 is associated with much stronger variance in the HadCM2 control simulation than in HAM3L or ECHAM1/LSG, while EOF 1 has only slightly more variance in HadCM2 than HAM3L (see Table 2)

simulations, but that correlations deteriorate if the spatial mean is not included in a correlation. The first EOF correlates also with the MPI G simulation and the HadCM2 G and GS simulations. The first, and to some extent the second common EOF, has counterparts in the first few EOFs from the individual models which have been used for truncation (see Appendix).

Table 2 Variance associated with common EOF patterns 1 to 4 compared between the control simulation of the three different coupled climate models $[(K/10yrs)^2]$

	EOF 1	EOF 2	EOF 3	EOF 4
HadCM2	0.99	0.98	0.37	0.23
HAM3L	0.86	0.17	0.24	0.20
HAM1L	0.18	0.29	0.45	0.49

Table 2 shows a comparison of the variance of the time evolution of the EOF patterns in each model's control run for the first few common EOFs. The strongest differences occurs for the first two EOFs. EOF 1 has strongest variability in HadCM2 and HAM3L, while EOF2 is entirely dominated by HadCM2. The first two EOF patterns resemble the common EOFs 2 and 1 of the analysis in Stouffer et al. (1999), where 5-year mean data of HadCM2, HAM3L and the GFDL model were studied in an otherwise similar analysis. It was found that HadCM2 has stronger tropical variability than most other models, and also seems to have stronger variability than the observations, while HAM3L showed too little tropical variability compared to the observations (note that a comparison with observations was not feasible on the 50-year time scale considered here). EOFs 3 and 4 show strongest variability in HAM1L. The results suggest that all three models show quite different partitioning of variance on 50-year time scales and that HadCM2 has more variance than both other models. In both HAM3L and HAM1L the variance is strongest in the NH landmasses (EOFS 1, 3, 4), while HadCM2 shows also quite strong variability in the Tropics (EOF 2). Since the covariance norm focuses on less variable areas (i.e., patterns associated with small λ_i) we expect the MPI fingerprint to yield a stronger focus on the tropics than the HAD fingerprint. This difference in focus is discussed in more detail in the Appendix.

5 Results

We have performed the optimal regression of the observations onto model signals using Had and MPI optimal fingerprints:

The MPI optimal fingerprint (see Eq. 3) is based on the HAM3L signal patterns, while the covariance matrix has been estimated from running 50-year summer trends from the first 900 years of the HAM3L control simulation (referred to in the following as "HAM-ctl1"; 851 trend patterns starting each year are used for the estimate). The results proved quite similar to the earlier detection and attribution results using 1000 years with HAM1L for estimating the covariance (H97).

The Had optimal fingerprint has been computed using HadCM2 signal patterns and 50-year summer trends from the first 900 years of the control simulation (Had-ctl1) to estimate the covariance matrix.

Amplitude estimates of anthropogenic greenhouse gas and sulfate signals can now be computed using both model's optimal fingerprints (Eq. 3). Overlapping 50-year summer trends from the control simulation that were not used for optimizing, i.e., from 810 remaining summers from the HadCM2 control simulation (Had-ctl2, year 901 to 1710) and 616 remaining summers from HAM3L (HAM-ctl2, years 901–1516) were then used for assessing the inherent variability in the signal amplitude estimates from observations and models due to internal climate variability (O1 and M1).

Note that the full 5° × 5° spatial pattern is of too high a dimension to reliably estimate its covariance of 50-year trends from even a control simulation of many centuries. Therefore it is necessary to truncate all data to a lower-dimensional base space, which allows for a reasonable representation of the signal patterns and of the most important climate noise features of each model's simulations. We have combined the approach described in Allen and Tett (1999) with some simple checks to determine the truncation level (10) for both models, this and other technical details are discussed in the Appendix.

5.1 Estimating a combined greenhouse gas and sulfate aerosol signal

Figure 4 shows the best estimate of the amplitude of a greenhouse gas and sulfate signal in observed and model simulated JJA trends over the same time period 1948–97. The uncertainty range for the observed GS signals originates from internal climate variability, whose variance is estimated by model internal variability (we show results from HadCM2, since it shows more variance than HAM3L in the GS amplitude and hence will yield conservative results, see also Fig. 6), and from observational sampling error (see Hegerl et al. 2000, for details). Both distributions are assumed to be Gaussian, their variances are added to the variance of total uncertainty in $\widehat{\boldsymbol{a}}_{obs}$. A Student-$t$ test assesses if the observed signal estimate $\widehat{\boldsymbol{a}}_{obs}$ is significantly different from zero, given the combined noise and sampling error uncertainty. The number of degrees of freedom (DOFs) in the noise samples from Had-ctl2 and HAM-ctl2 was estimated by the number of independent 50-year segments multiplied by 1.5 (21 for Had-ctl2, 16 for HAM-ctl2). The factor 1.5 is appropriate for estimates based on overlapping segments, provided the autocorrelation time scale is shorter than the segment length (Allen and Smith 1996) and has been also applied in T99 and S00. Earlier Monte-Carlo simulations (Hegerl et al. 1996) simulating correlated samples with an AR-1 model suggested that this is a rather conservative approach. Results were insensitive to reasonable variations in estimated DOFs. The confidence bars shown in Fig. 4 indicate the 90% uncertainty range based on the Student-t critical values. The uncertainty range does not include the case of no signal (zero), independent of which model's GS fingerprint is used. Therefore we conclude that the recent 50-year trend patterns contain a climate change signal with significance level of 5% (1-sided test).

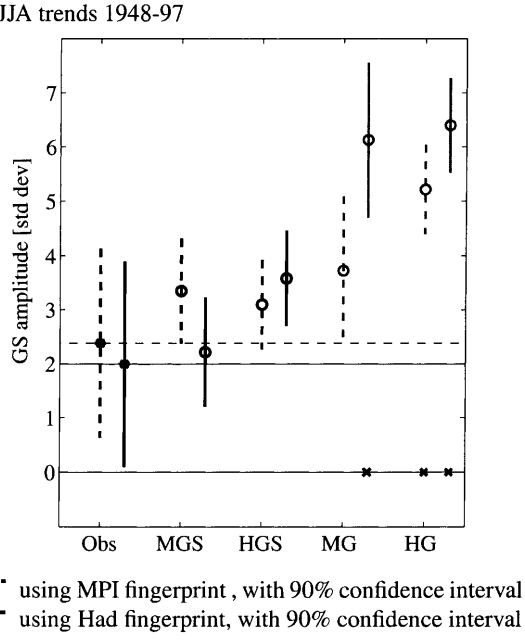

Fig. 4 Comparison of the amplitude of a single anthropogenic GS signal in observations and model data in the JJA surface temperature trend pattern 1948–97 (for comparison, the amplitudes are renormalized to standard deviations of HadCM2 internal variability). The amplitude estimates based on the MPI GS fingerprint (based on the signal shown in Fig. 2a) and that of the Had GS fingerprint (Fig. 2b) are both shown. *Obs* denotes the observed trend pattern, *MGS* that of the HAM3L GS simulation, *HGS* that of the HadCM2 GS simulation, and *MG* and *HG* the respective greenhouse gas only simulations. Simulations which are inconsistent with observations are marked by a *cross* on the *horizontal axis*. A significant GS signal is detected in observations with either fingerprint

The 90% uncertainty bars for the amplitude estimates from the model simulations are also computed from critical values of a Student-t test. The noise variance in that case is based on the variance of the model control simulation divided by the size of the ensemble which has been averaged (2 and 1 simulations for HAM3L, 4 for the Had simulations). All model signals are significantly positive. Estimating one model's signal amplitude using another model's fingerprint can be interpreted as a model-model study of the potential impact of model

error on the estimation of signals, results indicate that the effect of model difference is visible, but limited. To determine if the difference between observed and simulated amplitude estimates $\widehat{\boldsymbol{a}}_{obs} - \widehat{\boldsymbol{a}}_{M}$ (using the same fingerprint) is significant, the difference is compared with the sum of the variances of noise in observations O1 and model data M1, and sampling error O2. A Student-t test yields that only the trend patterns from the greenhouse gas only simulations give unrealistic GS amplitudes.

5.2 Separating a greenhouse gas and a sulfate aerosol signal

Using several fingerprints can help to distinguish between climate response to different forcing and hence to attribute anthropogenic climate change to causes (Hasselmann 1997, H97; T99). We therefore attempt to estimate a greenhouse gas only type signal (G) and an orthogonal sulfate signal (Sul). Sul is computed by orthonormalizing the GS signal pattern relative to G (Eq. 2). Note that the Sul pattern is not the temperature response to sulfate aerosol forcing only, which produces global mean cooling in addition to the Sul pattern and is anticorrelated with a greenhouse gas pattern. This "global cooling" component of the sulfate signal will reduce the estimated G amplitude in our analysis, whereas the Sul signal reflects purely the pattern component which is unique for sulfate aerosol forcing. Figure 5 shows Sul patterns from both models. Both patterns show strong NH midlatitude cooling as the most important pattern difference to G. However, the MPI Sul pattern shows substantially more warming in the SH outer tropics and lower mid-latitudes.

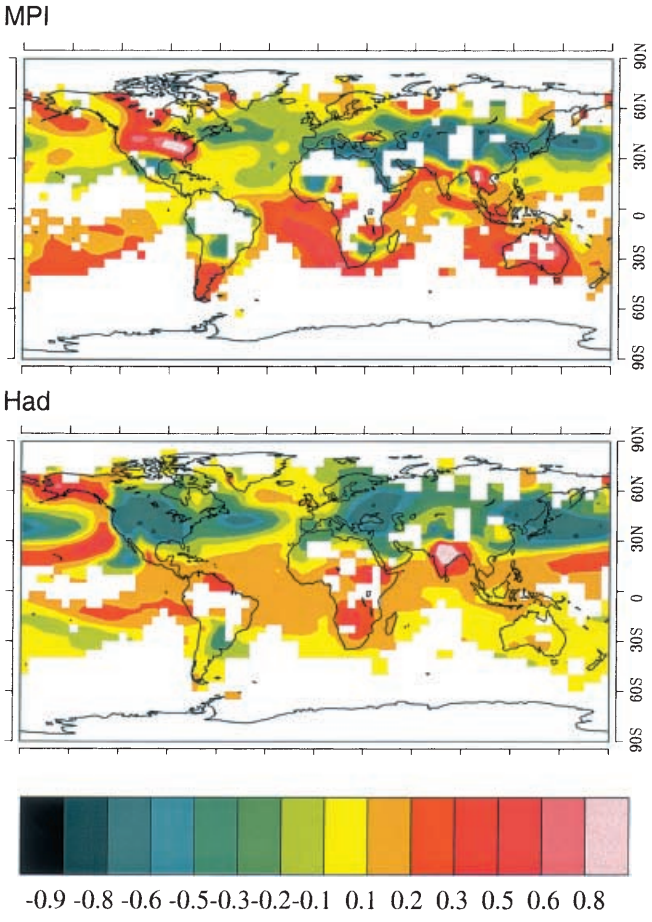

Fig. 5 Orthogonal sulfate aerosol pattern in both model simulations. This pattern is derived by orthogonalizing the GS pattern to the G pattern for each model (see Figs. 1 and 2) and shows the pattern influence of sulfate aerosol forcing which is dissimilar to that of greenhouse gas forcing for each model. Note that the Had orthogonal sulfate pattern resembles common EOF 1 more than the MPI orthogonal pattern (see Figure 3, Table 1), and is therefore more difficult to separate from climate variability

Figure 6 shows the results of the attribution analysis using both Had and MPI fingerprints. The results show that the latest observed pattern of linear trends between 1948–1997 contains both a positive greenhouse gas signal (value along the horizontal axis) and a positive sulfate signal (value along the vertical axis). The magnitude of both signals varies depending on whether the Had or MPI fingerprints were used. The combination of greenhouse gas (G) and sulfate signal (Sul) which composes the GS signal pattern in each model is shown by the slanted labelled line (see H97 for details). Note that one-signal GS detection results shown in Fig. 4 can be obtained by projecting all data onto that GS axis. The much smaller angle between the GS and the G only axis using the Had fingerprints rather than that using the MPI fingerprints relates to the greater pattern similarity between both signal patterns for HadCM2, which has been discussed in the previous section.

As for the single-signal detection, we have computed amplitude variations from the HAM-ctl2 and Had-ctl2 control simulations to estimate variations in both signal amplitudes due to climate noise. The ellipse around the observations gives the 90% uncertainty range of internal model noise and sampling error. As in the single signal case, the variance of noise and sampling error have been added and a Hotelling-T^2 test has been performed to determine critical values for the null hypothesis that the true observed signal amplitudes are both zero (see von Storch and Zwiers 1999). The uncertainty ellipses for the model simulations are based on the variance of each model's noise divided by the number of ensemble simulations (using Hotelling-T^2 critical values). Note that the reliability of the assumption of Gaussian noise is discussed in H97, the data appeared to be quite consistent with this assumption. In H97, a test assuming well-sampled Gaussian noise was performed, taking sampling uncertainty into account in the present work by applying critical values for a Hotelling-T^2 test increased the ellipses only slightly.

Applying the MPI fingerprints, amplitudes of zero for both the G and the Sul pattern in the observations are rejected in a two-dimensional and in one-dimensional tests for each signal (at a less than 5% risk level, Table 3). On the other hand, while a G signal can be

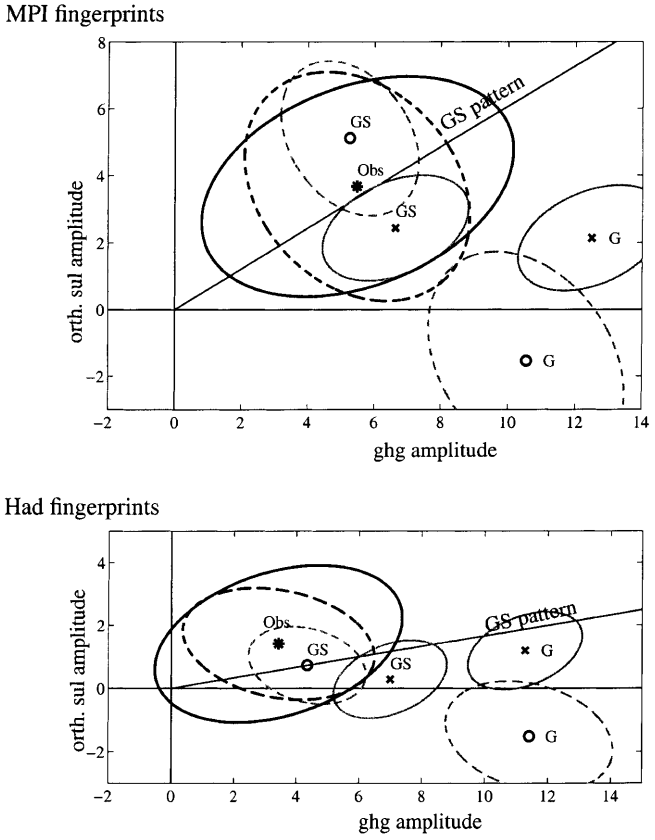

Fig. 6 Comparison between estimates of the greenhouse gas and orthogonal sulfate signal in present observed 50-year summer trends 1948–1997 (*black star*) and model simulations for G and GS climate change. *Horizontal axis*: amplitude of a G climate change signal; *vertical axis*: amplitude of the orthogonal sulfate signal and skewed axis: amplitude of a GS signal (normalized according to Eq. 2). The results shown in the *top panel* are based on MPI optimal fingerprints, those in the *bottom panel* on Had optimal fingerprints. The *ellipses* shown in *black* are a 90% uncertainty ellipse for the observed signal due to internal climate variability (estimated from HadCM2, *solid*, and HAM3L, *dashed*) and sampling error. The signal amplitudes from JJA trend patterns 1948–97 in model simulations are shown by *crosses* for HadCM2 and by *circles* for HAM3L. Each model value is subject to uncertainty due to superimposed model noise, for which 90% uncertainty *ellipses* are shown in *grey*. The size of the ellipse depends on the size of the ensemble (G, HAM3L: 1, GS, HAM3L: 2, G and GS HadCM2: 4). A one-dimensional test indicates that the observations contain a significant G climate change signal; if MPI fingerprints are used, they also a significant orthogonal sulfate signal. Significance levels for the difference between each model simulation and the observations are given in Table 3. Each model's greenhouse gas only simulations show highly significant disagreement from the observations. Using the MPI fingerprints the present trend in both model's GS simulations is not inconsistent with the observations. Using the Had fingerprints, the Had GS simulation disagrees from the observations at a 10% significance level

detected with Had fingerprints, the joint detection of G and Sul includes the possibility that both are zero. Note that rejecting the null-hypothesis that both signals are zero in a two-dimensional test is a more demanding objective (requiring a greater absolute discrepancy for rejection) than testing whether *either* is equal to zero (for a discussion see Hasselmann 1997). The amplitude of the Sul signal in observations is estimated as substantially smaller relative to climate noise if the Had optimal fingerprints are applied and is not significant. This difference to the results based on MPI fingerprints can be better understood if the spatial patterns of the orthogonal sulfate aerosol pattern (Fig. 5) are analyzed. Climate variability projects stronger on the Had Sul pattern than on the MPI Sul pattern (as indicated by a pattern correlation of about −0.8 between Had Sul and EOF 1 contrasting with −0.2 between MPI Sul and EOF 1, Table 1). Therefore, the sulfate aerosol influence is less separable from climate noise using Had fingerprints. We conclude that the sulfate influence on observed trend patterns cannot be reliably estimated from temperature trend patterns using the orthogonal sulfate fingerprint, since the estimate is found to be model dependent. Note that T99 and S00 found both a significant G and Sul signal in their time-space analysis, suggesting that the time-component helps to distinguish a sulfate signal.

To determine if the difference between observed and simulated amplitude estimates is significant, the Hotelling-T^2 test is performed with the null-hypothesis that the "true" difference is zero (analogous to the one-dimensional Student-t test). The results show that both model's G only simulations are highly and significantly different from the observations, due to a too strong amplitude of the greenhouse gas signal in the simulations (indicating too strong global mean warming), note that the discrepancy is now more obvious than from the single signal analysis (Fig. 4).

A significant disagreement at the 10% level is also found between the Had GS simulation and observations if Had optimal fingerprints are applied (Table 3). The disagreement is due to both a too high amplitude of the greenhouse gas signal pattern and a too small amplitude of the simulation's sulfate pattern. Regardless of whether Had or MPI optimal fingerprints are used, no significant disagreement between the MPI GS simulation and the observations is found. A comparison of trend patterns for 1948–1997 between the observations (Fig. 7, bottom) and both model's GS simulations (Fig. 8) shows that the HadCM2 simulation shows more warming in the NH midlatitudes than either the observations or the HAM3L simulation. No significant disagreement is diagnosed for HadCM2 trend patterns ending in 1995 (see Fig. 11, bottom). A recent analysis of an updated observed dataset also showed no disagreement for trend patterns ending in 1998. (Results for 1998 yield a significant joint detection of G and Sul, but otherwise very similar results, Hegerl et al. 2000). Therefore it is uncertain if this small disagreement relates to an unrealistic aspect of both model's forcings or in the HadCM2 response, or if this disagreement is due to a 10% rare noise event in the HadCM2 simulated trends ending in 1997.

We conclude from the one- and two signal analysis that anthropogenic combined greenhouse gas and sulfate forcing is a possible reason for the observed climate change. In both models the additional sulfate aerosol

Table 3 Results of testing the null hypothesis that the observations agree with climate noise (obs = 0) and various climate change simulations

Optimal fingerprints Observed noise estimate	MPI HAM3L	MPI HadCM	Had HadCM	Had HAM3L
Obs = 0 (2-D)	1%	5%	–	1%
Obs G = 0	1%	1%	5%	5%
Obs Sul = 0	5%	5%	–	5%
Obs = G (MPI simulation)	5%	5%	1%	1%
Obs = G (Had simulation)	1%	1%	1%	1%
Obs = GS (MPI simulation)	–	–	–	–
Obs = GS (Had simulation)	–	–	10%	10%
Obs = Nat (Had simulation 1946–95)	5%	10%	10%	10%
Obs = GS + Nat (Had simulation 1946–95)	–	–	–	–

If the null-hypothesis is rejected, significance levels are given, non-rejection is indicated by a "–". The first two columns give results based on the MPI fingerprints (using HAM3L or HadCM2 to estimate observed climate noise); the second two show results based on Had fingerprints. Hotelling-T^2 Tests have been applied for 10%, 5% and 1% significance levels, details see text. Obs G = 0 and Obs Sul = 0 refers to a *t*-test of the null hypothesis that the observations contain no G or no Sul signal. For naturally forced simulations, which end in 1995 (HadCM2 Sol and Vol, Sol + Vol is indicated as Nat), the trend patterns 1946–1995 are compared with contemporaneous observations. Note that the results based on MPI fingerprints are very similar to those in H97, where HAM1L was used for optimization

forcing causes an overall reduction in warming trends which yields a better agreement of these simulations with observations. The results were found to be robust to truncation decisions (i.e., the dimension of the space chosen for representing model and observational data), using the HAM1L rather than HAM3L control simulation for computing MPI optimal fingerprints as in H97 (not shown), and to swapping the data used for estimating the covariance with those used to estimate the internal variability (using Had-ctl2 and HAM-ctl2 to compute optimal fingerprints, and Had-ctl1 and HAM-ctl1 to estimate variations in amplitude estimates due to internal climate variability only).

The difference between 1948–97 trends from the two model's simulations in Fig. 6 was found to be statistically significant at the 10% level based on a Hotelling-T^2 test, for both the G and GS simulations. The future evolution of the amplitude estimate from the climate change simulations was found to be increasingly different: while the HAM3L GS and G only simulations separate strongly in the Sul direction, (indicating very large positive hemispheric asymmetry for GS if MPI fingerprints were used), no such separation was found for the HadCM2 simulations (not shown). The physical mechanisms responsible for such a difference are beyond the scope of this work, we refer the reader to the studies of Räisänen (1997, 1999).

Recently, model simulations based on several other, newer simulations, some of them with different sulfate forcing, have been added in an attribution test similar to that shown in Figs. 4 and 6 (Barnett et al. 1999). Not all model's GS simulations were found to agree with observations, further supporting the importance of model and forcing error in detection and attribution approaches.

5.3 Optimal model-model comparison

The smaller estimates of both climate change signals using the Had fingerprints, and the better separation of the climate change signal from noise using the MPI fingerprints suggest that the MPI fingerprints filter the signal more effectively from the noise than the Had fingerprints. To further investigate this issue, we have synchronously regressed the observations on both models' GS patterns (using again the inverse covariance inner product; Eq. 1). Such an exercise investigates whether the observations show more agreement with one model's signal pattern than with the other. To be fair to each model, the computation of the covariance has been done using both MPI and Had data. We show results using the Had-ctl1 estimate of the covariance, the results using MPI data were quite similar.

As in the case of the G and orthogonal Sul fingerprints in Fig. 6, we projected the observations onto the Had GS signal pattern and that component of the MPI GS pattern which is orthogonal to it relative to the inverse covariance as inner product. The orthogonal signal pattern (Fig. 2c) shows that the most important differences are relatively less warming in the tropics in the HAM3L pattern, more warming in the SH, and more warming over parts of the midlatitude oceans and Southern North America.

Figure 9 shows the results. The horizontal axis gives the amplitude of the Had GS pattern, while the skewed axis shows the amplitude of the MPI GS pattern. The relatively large angle between both is caused by the relatively large difference between both signal patterns, as has been discussed previously. The recent observed trend pattern is rather close to the MPI GS pattern (indicated by the skewed line), also, the amplitude of the orthogonal MPI pattern in the observations is significantly different from zero (significance level $< 5\%$). This indicates that the observations contain a pattern which more closely resembles the HAM3L signal pattern than the HadCM2 signal pattern. Zero amplitude of both GS signals is excluded at a very high significance level ($< 1\%$ for both 1-dimensional and 2-dimensional test results), showing that the information from both models' GS fingerprints combined leads to a particularly

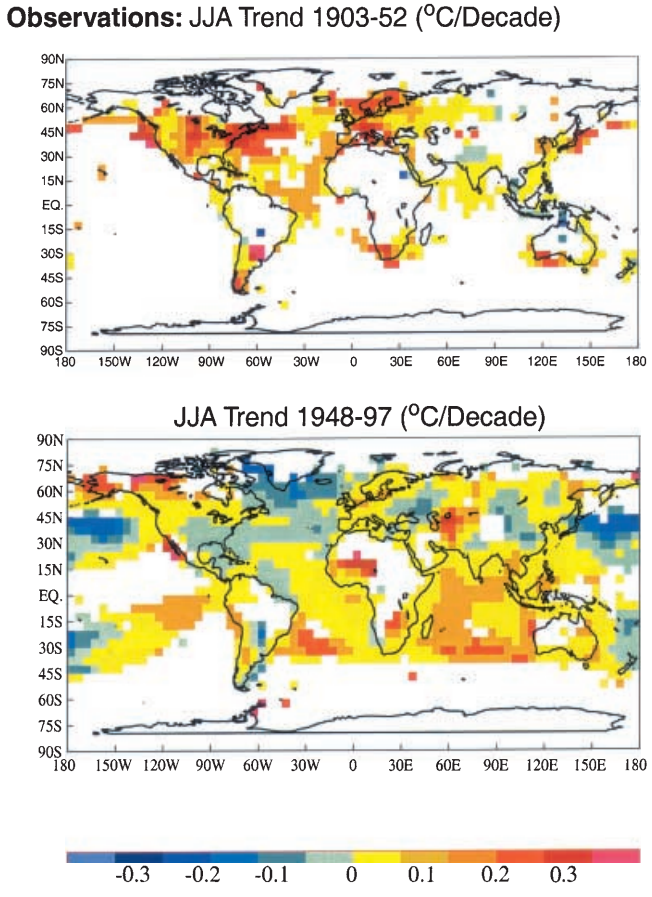

Fig. 7 Comparison of observed trend patterns of Northern Hemisphere summer (JJA) between 1903–1952 (peak of the early century warming) and 1948–1997. The pattern 1946–1995 (not shown) is similar to the lower panel

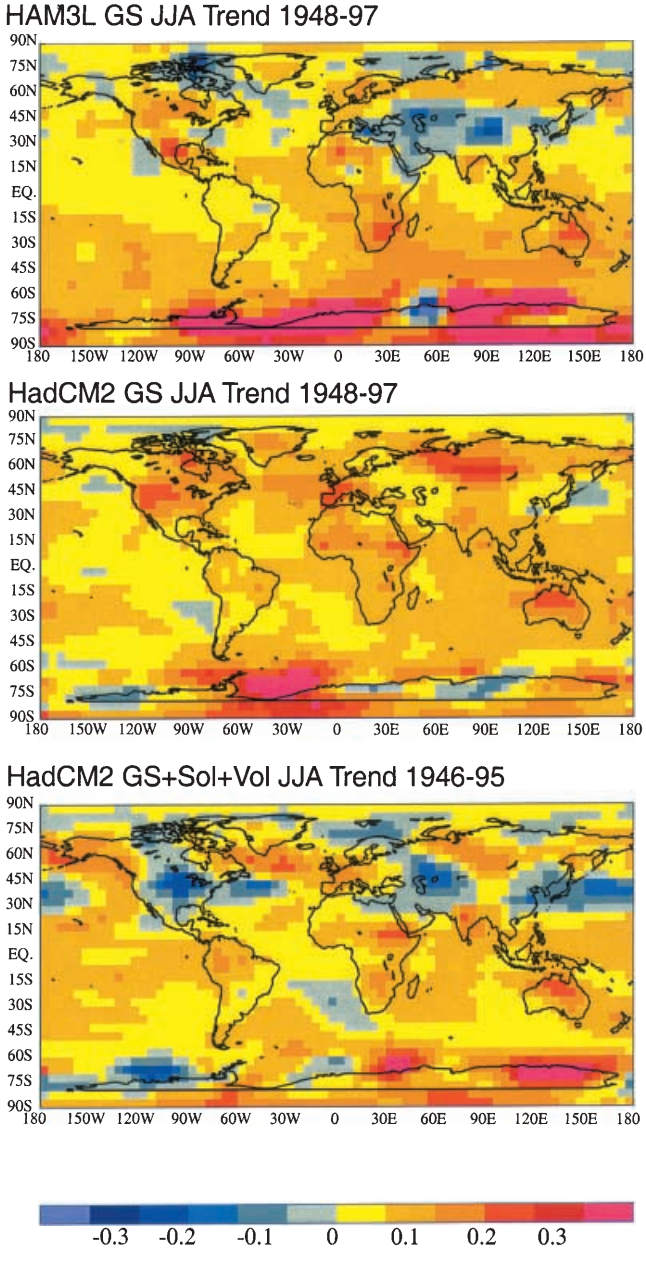

Fig. 8 Same as Fig. 7, but for simulated trend patterns. *Top panel*: HAM3L GS simulation, *middle panel*: HadCM2 simulation and *bottom panel*: linear combination of HadCM2 simulations forced with greenhouse gases, sulfate aerosols, solar and volcanic forcing (trends ending in 1995). Note that the trend patterns differ from the GS signal pattern (Fig. 2) due to superimposed climate noise and, possibly, due to time-varying aspects of the model response

clear detection of climate change. Figure 9 also shows that using both model fingerprints in parallel is a very promising method to isolate model differences. In this case, no significant difference of the observed trend pattern 1948–97 with either model's trend pattern could be isolated, since the observations are located diplomatically between trend patterns from both models.

5.4 The potential influence of natural climate forcings

To assess how estimates of anthropogenic signals depend on the time period which was analyzed, Fig. 10 shows the time evolution of the greenhouse gas signal and the orthogonal sulfate signal from running observed trend patterns. For a better comparison of results based on MPI and Had fingerprints, all amplitude estimates have been re-normalized to standard deviations of the second chunk Had-ctl2 of the HadCM2 control simulation (which has not been used to compute the optimal fingerprint; note that it shows larger noise than HAM3L). Overall, the qualitative impression of the time evolution of the anthropogenic climate change signal agrees between using different fingerprints. Similar to the most recent trends, the observed sulfate signal is estimated as consistently smaller when the Had Sul fingerprint is used than when the MPI fingerprint is used. The estimated Sul signal amplitudes increase with time independent which model is used.

A rather striking feature of Fig. 10 is that the estimate of the anthropogenic greenhouse gas only fingerprint is larger during the early century warming (between ca 5.2 to 5.7 standard deviations, SD) than at

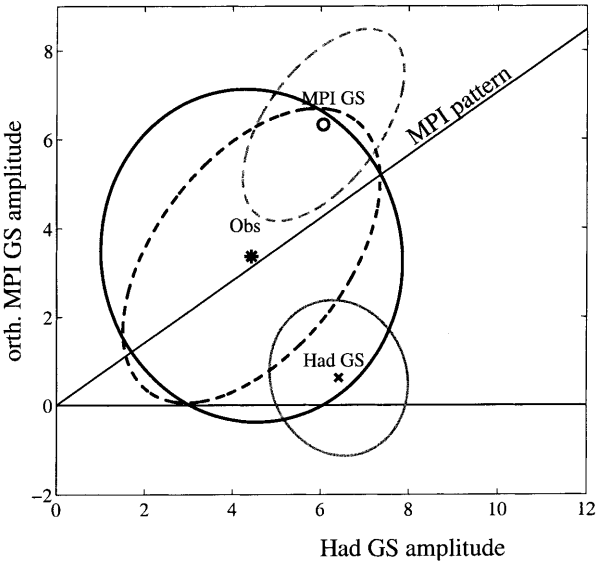

Fig. 9 Comparison between the recent observed summer trends 1948–1997 and those from GS simulations, applying both model's GS signal patterns simultaneously. This approach focuses on the model related differences in that signal pattern. *Horizontal*: amplitude of the Had GS pattern; *slanted line*: amplitude of the MPI pattern, *vertical*: pattern difference (see Fig. 2c). The 90% confidence intervals for the observed and model trend patterns are shown, see Fig. 6 caption. The results suggest that both models show marginally unrealistic features, and that the observations agree better with the GS signal pattern from HAM3L. Note that a highly significant climate change is detected

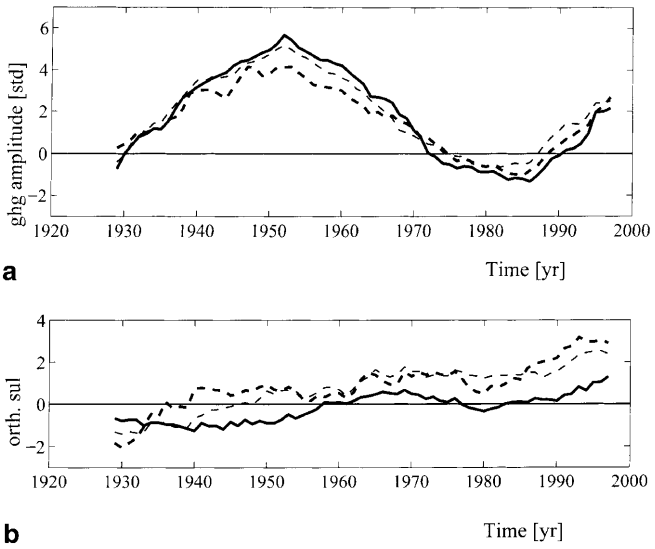

Fig. 10a, b Evolution of the amplitude of **a** a greenhouse gas-only and **b** the orthogonal sulfate signal from running observed trend patterns, starting with the trend 1880–1929 (plotted at 1929) and ending with the recent trend (*solid line*: estimate based on the Had fingerprints, *dashed line*: MPI fingerprints, *thin dashed line*: MPI fingerprints optimized with HAM1L). All amplitude estimates have been rescaled to units of standard deviations of the HadCM2 Control simulation, the year denotes the last year of the observed 50-year trend period used to estimate the signal amplitudes. The greenhouse gas only signal amplitude is strongest in the period associated with the early twentieth century warming (possible reasons see Fig. 11)

the present time (2.15 to 2.5 SD). The largest estimate of the greenhouse gas signal amplitude during the early century warming originates from the Had fingerprints, while the amplitude of the present warming is estimated larger from the MPI fingerprints. This difference should be due to the difference in both model's optimal fingerprints, which emphasize different components of the observations in that period: the summertime trend pattern between 1903–52 shows an overall strong warming in almost all areas covered by observations, peaking in the Atlantic Ocean and the adjacent continental regions, while the most recent warming is quite large in the tropics and the SH oceans (Fig. 7). We speculate that the early century warming in the Northern Hemisphere will be partly suppressed by MPI fingerprints, since this model has strong variability in that region (common EOF 1, 3, 4; Eq. 4). Conversely, HadCM2 will suppress more of the recent tropical warming (EOF 2). This is discussed more in the Appendix. Note that the high amplitude estimates of an anthropogenic signal in the first half of the twentieth century disappears entirely if spatial mean warming is disregarded in optimal fingerprint results, while the recent amplitudes are not sensitive to disregarding global mean warming (Hegerl et al. 2000). This agrees with the observations showing a rather dissimilar pattern to the G signal pattern in the first half of the twentieth century (Fig. 7), suggesting that the high G amplitude originates at least partly from sources other than greenhouse gas forcing.

One conceivable explanation for the higher greenhouse signal amplitude during the early twentieth century is that it represents climate noise only. If this were the case, it would imply that it was either a very unusual event, or that both models seriously underestimate climate variability, compromizing detection results. To investigate whether the early twentieth century warming is a typical fluctuation associated with natural climate variability, 50-year trends have been computed from NH summer temperature reconstruction since 1000 AD (Jones et al. 1998). The trend ending in 1952 is the largest trend in the entire proxy time series (not shown), showing that the early century warming was a truly unusual event (once in a 1000 year proxy-time series, and never in the model's control simulations). While this does not prove that model variability is correct, it is reassuring that paleo data support the impression that the early twentieth century warming is not a typical expression of internal climate variability.

To investigate whether external climate forcing might have influenced this exceptional warming the optimal detection and attribution analysis has been repeated for the trends 1903–1952, excluding all gridpoints from the analysis which were not covered by observations then (see Fig. 7). We study anthropogenic and naturally forced climate simulations. Since simulations forced with both anthropogenic and natural forcings are not available, we have linearly added simulations forced with anthropogenic, solar and volcanic forcing. Note

that a linear superposition of model simulations disregards nonlinear interactions between different forcings and also increases the uncertainty level, since the variance of noise in each simulation is additive. Also, our limited knowledge of natural forcing on climate by volcanism and solar radiation changes makes these results very uncertain.

Since a meaningful statistical test cannot be performed on a particularly unusual, but unexplained feature in the observations (which may be, after all, a hand-picked extreme event), the observation is surrounded by amplitude fluctuations associated with climate noise from the control simulations indicating where the true signal might lie given superimposed climate noise. Both model's GS simulations are found to show far smaller signals in 1903–52 than the observations. The difference is larger than can be explained by adding the most extreme event in either control simulation. Note that this may be model-dependent, Delworth and Knutson (2000) recently found that one of five GS realizations with the GFDL model shows very similar warming in the first half of the twentieth century as observed. The HAM3L and HadCM2 greenhouse-gas-only simulations (not shown) yield hardly more warming than the GS simulations. Solar forcing alone (as simulated in HadCM2 and HAM3L's Sol simulations) performs similarly poorly in explaining the early century warming. A linear superposition of solar and volcanic simulations (Nat) together with an extreme noise event might agree with the observations. If Nat is superimposed on the anthropogenic (HadCM2) GS simulation, the results appear quite consistent with the early twentieth century warming.

This result suggests that natural forcings project onto estimates of anthropogenic greenhouse gas and sulfate signals. This complicates interpretation of the attribution test results in Fig. 6, which considered anthropogenic forcing only. We therefore have repeated the attribution test for the observed trends 1946–1995 (when the Sol and Vol simulations end) taking natural climate forcings into account. The results are summarized in Fig. 11 and Table 3. Naturally forced climate change alone (Nat) is found to be inconsistent with the observations at present, since it would yield a slightly negative greenhouse gas amplitude. A solar forced simulation with the MPI model is only rejected if the old HAM1L fingerprints are used (at the 10% significance level, see also H97), while that with HadCM2 is not rejected. However, the same HadCM2 solar forcing simulations were found to be inconsistent with the observations in T99 and also S00 (where optimal time-space detection was performed). The superposition of solar, volcanic and anthropogenic greenhouse gas-plus-sulfate forcing is found to agree best with the observations, which is confirmed by visual inspection of the trend pattern (Fig. 8).

This result is consistent with the findings in T99, who found a significant greenhouse gas and sulfate signal, particularly in the most recent 50-year time period. S99 detect a solar signal in the early twentieth century,

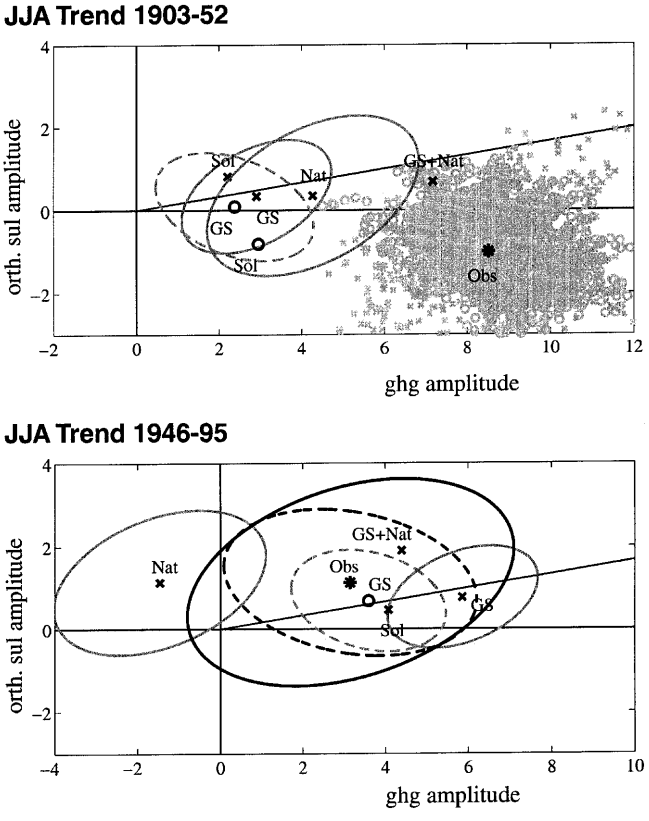

Fig. 11 *Top*: comparison between the observed summer trends 1903–52 (*star*) and the respective simulated trends for greenhouse gas + sulfate aerosol climate change, solar forcing (*Sol*), a superposition of solar and volcanic forcing (*Nat*), and the superposition of anthropogenic and natural forcings, in an attribution diagram similar to Fig. 6. Results are shown based on the Had fingerprints. Typical 90% confidence ellipses for HadCM2 (through) and HAM3L simulations (*dashed*) are shown. The observations are surrounded by all samples of internal climate variability of both model's control simulations ("+" for HadCM2, "o" for HAM3L, both in *grey*), indicating the possible spread of the true underlying signal minus even extreme noise events. The results show that unless natural climate forcings are considered, only extremely rare events of internal climate variability could explain the early twentieth century warming given the HadCM2 and HAM3L GS simulations. *Bottom*: attribution diagram for the summer trends 1946–95 for natural and anthropogenic forcings. The results indicate that natural and anthropogenic forcings combined agree between HadCM2 model simulations and observations also for the present, while natural forcings alone disagree significantly

however, only if based on one of two solar forcing reconstructions (the same used in the present work). Both T99 and S99 rule out solar forcing alone as a cause of the present warming. The detection of a volcanic signal in S99 was dependent on which decades were analyzed and remains uncertain.

We conclude that a combination of natural and anthropogenic forcings agrees quite well with the observations during the early twentieth century and also at present (the time evolution of the signal amplitudes suggests an agreement over most of the twentieth century, not shown). A combination of natural (solar and volcanic) forcings alone appears an unlikely explanation of observed climate change and has been rejected in the

present attribution test and in T99 and S00. However, these results are subject to uncertainty in our limited knowledge of the effect of natural forcing on climate.

6 Conclusion

The optimal detection and attribution approach of Hasselmann (1997) has been applied to observed summer (JJA) trends over 50 years, using optimal fingerprints based on data from two different OAGCMs. This allows for a qualitative estimate of the impact of climate model uncertainty on detection and attribution results.

Significant differences between each model's response to the greenhouse gas and sulfate aerosol forcing have been found, although the global mean surface temperature response is similar (Kattenberg et al. 1996). In particular, the differences between the signal patterns due to greenhouse gas only and greenhouse gas-plus-sulfate aerosol forcing are larger in HAM3L than in HadCM2. One reason is that the Hadley Centre greenhouse gas only simulation contains pattern components which resemble the HAM3L aerosol signal, for example a reduced warming over the NH midlatitude oceans due to a relatively strong cloud feedback and less pronounced warming in the NH midlatitudes. Also, HadCM2 simulations show strong warming in the tropics for both greenhouse gas only and greenhouse gas-plus-sulfate aerosol forcing.

Additionally, an analysis of internal climate variability on 50-year time scales reveals differences between the HadCM2 model and two MPI models (ECHAM3/LSG and ECHAM1/LSG). HadCM2 shows large tropical variability, whereas in both MPI models, the largest variability occurs over the Northern Hemispheric landmasses. These differences influence detection and attribution results. In particular:

1. A sulfate aerosol-induced pattern alone can be detected in the observations using the MPI fingerprints, while the sulfate response using Had fingerprints is not separable from a reduction of the greenhouse gas amplitude. This appears to be partly due to a stronger hemispherically asymmetric warming in the MPI GS signal pattern than the Had GS signal. The simultaneous detection of greenhouse gas and sulfate aerosol patterns in observations is only possible with MPI fingerprints (note, however, that this is a very stringent test and not necessary for attribution of climate change to human activity).

2. A recent significant climate change is filtered more effectively from the data using the MPI fingerprints than the Had fingerprints. This is due to the fact that the observations more closely resemble the HAM3L greenhouse gas-plus-aerosol pattern than the respective HadCM2 pattern. This result, however, does not necessarily mean that the HAM3L model is more realistic. An alternative explanation would be that errors in the forcing history of both models or missing forcings are partly cancelled by errors in HAM3L's response.

These results show that model uncertainty has an important bearing on attempts to quantify anthropogenic and natural climate change signals in the observations. The following conclusions appear robust to model uncertainty:

1. We have high confidence that the recent temperature evolution is inconsistent with estimates of internal climate variability alone. This result was particularly stable in earlier work based on 30-year annual mean trends, and can also be clearly demonstrated for the most recent 50-year summer temperature trends. Similar results are obtained using the HadCM2 data if time-space optimal detection is applied (Tett et al. 1999; Stott et al. 2000). Note that model uncertainty is not as important for detection of climate change as for attribution of that change to causes, since errors in model signals will lead to an underestimate of a signal. The largest caveat here lies in the estimates of climate variability.

2. Greenhouse gas forcing alone yields a significant difference between the observations and both model simulations. A combination of anthropogenic greenhouse gas and sulfate aerosol forcing (GS) is a more probable explanation, the aerosol influence in both models causes a delay in the warming which yields better agreement of combined greenhouse gas plus sulfate aerosol simulations with present temperature trends than simulations forced with greenhouse gases alone. The HadCM2 simulation shows marginally significant differences to one of three of the recent observed trend patterns, indicating that HadCM2 may show slightly too strong warming.

3. There is suggestive evidence that a combination of natural and anthropogenic forcings has influenced the twentieth century climate evolution. The early century warming was a very unusual event in a 1000 year proxy record. Results indicate that it was either caused by a superposition of anthropogenic climate change and a very extreme event of internal climate variability, or is influenced by additional natural external forcing. The agreement between observed and (HadCM2) simulated temperature trends improves for both the early twentieth century and the present period if volcanic and solar forcing are taken into account. On the other hand, natural climate forcings alone are found to be insufficient to explain the twentieth century temperature evolution. These results agree qualitatively with Tett et al. (1999) and are particularly sensitive to uncertainties in external climate forcing (note in particular that, with the exception of greenhouse gases, the forcings are highly uncertain) and in the climate model's response to them.

The results highlight a need to understand the physical mechanisms causing the differences in climate model's response to external forcing and the reasons for differences in their internal variability in order to yield more reliable estimates of the magnitude of the various external climate forcing responses in the observations. Also, a reduction in the forcing uncertainties is required.

Acknowledgements The authors wish to thank Klaus Hasselmann, Reiner Schnur, William Ingram and Reinhard Voss for support and valuable discussions, Ben Santer for a very thorough review with many constructive suggestions, and G. North, T. Crowley, B. Hyde and D. Ritson for comments. We are grateful to P. Jones for the latest near-surface temperature data and the NH paleo reconstruction. Different components of the work were sponsored by (for GH:) The Alexander von Humboldt-Stiftung, JISAO (contribution 645) NSF (grant ATM-9707069), and NOAA's Office of Global Program's Climate Change Data and Detection Program Element, and DOE's Office of Health and Environmental Research the Hadley Centre under UK-DETR and UK-NERC (SFBT, MA, PAS, JFBM), by UK Department of the Environment, Transport and the Regions under contract PECD7/12/37 (PAS and SFBT), by the UK Public Meteorological Service Research and Development program (JFBM) and by the EC project SIDDACLICH (UC).

Appendix: sensitivity of results to technical details

Construction of the signal pattern

The GS and G signal patterns here have been isolated by an EOF analysis of the average of each model's ensemble simulations, using only the first EOF in each case. In the following details of the construction and the use of higher EOFs are briefly discussed. Since the ensemble size is larger for HadCM2, we have used a shorter base period (1935–2034) for the EOF analysis, compared to 1880 to 2049 for HAM3L, in the hope of providing a more exact representation of the present climate change pattern while reducing noise similarly as for the HAM3L simulations. However, earlier results with HAM3L indicated that the result of an EOF analysis was not particularly sensitive to the exact choice of analysis period.

In order to assess whether higher EOFs contain information about the present climate change signal, trends over 50 years have been computed from the first five EOF's time evolution between 1935 and 2034 in both models. Only the first EOFs in the G and GS simulations in both models showed trends which were convincingly different from climate noise. Since there was some indication that the second EOF in the GS simulations, particularly in HadCM2, might also contain a signal, sensitivity tests have been performed composing the signal pattern according to the magnitude of the trends in their principal components at the "present" time. The results were found to be in general similar to the results using only the first EOF. If the first two EOFs were used to compose the Had signal pattern, the similarity between the GS and the G signal was even larger than if only the first EOFs were used, making the detection of a Sul signal more difficult and the Had GS fingerprint more dissimilar to the observations in the model comparison (Fig. 7).

Choice of truncation level and effect of optimization

The estimate of the covariance of climate noise is vital for optimal detection and attribution (Eqs. 2, 3). Since the covariance of climate noise at the 50-year time scale is not known and cannot be estimated from the instrumental record, long model control simulations are used. However, estimates of the covariance matrix of multi-decadal trends in the gridpoint space will be degenerate even if estimated from 900 years of data. Therefore, all data used in the detection and attribution approach need to be truncated to some low-dimensional space (see Bell 1986; Hasselmann 1993) which allows for a good representation of the signal patterns and climate noise. We truncate to the EOFs of each model's concatenated GS simulations (the simulations have been concatenated rather than averaged to resolve signal and noise, H97). Sensitivity tests yielded similar results if the EOFs of the G simulations were used.

Care needs to be taken in determining the truncation level since a too high truncation level will yield erroneous results (Hegerl et al. 1996; Allen and Tett 1999; Leroy 1998) by too strong a focus on poorly sampled components of the covariance which will have underestimated eigenvalues, or on unrealistic features of model variability (e.g., associated with small spatial scales). Too low truncation, on the other hand, will not allow us to use the full power of the optimal detection method. To determine the truncation level n we have applied a χ^2 test as proposed in Allen and Tett (1999). The null hypothesis is tested that the observed noise in the residual $\tilde{\Psi}^r$ (Eq. 1) is smaller than or equal to that in the climate model in the same $n-p$ dimensional subspace (with p denoting the number of signals considered, Eq. 1). Therefore this test will not only determine if the optimal detection focuses on aspects of model variability which are unrealistic due to sampling problems (in which case the observed residual will become larger than the subsampled model control component, Eq. 4), but also if unreliably low noise influences some components of the optimal fingerprints. Note that if the overall noise level of a model is high, the null hypothesis will not be rejected even if an increase in truncation level n starts focusing on unrealistic or undersampled features. Therefore the evolution of the χ^2 values with n was considered in addition, with a sudden trend in χ^2 values indicating that results start to deteriorate $n = 10$ was found to be appropriate and safe for both models.

The **MPI optimal fingerprint** has been computed after truncation of all data to the first 9 EOFs of the GS simulation, augmented by the G signal pattern. This space allowed for a quite good representation of the signal patterns (on the order of $r = 0.99$ correlation between the original and the truncated patterns). The χ^2 test yielded non-rejection of the Null-hypothesis for all truncation levels considered (for technical reasons, only a truncation level up to 20 was possible). However, a truncation level beyond 15 was found to yield increasing disagreement between the noise levels of the HAM-ctl1 (which was used to estimate the noise covariance) and HAM-ctl2, suggesting focus on subsampled variability. In H97, a truncation level of 6 was applied, using the earlier HAM1L control simulation for estimating the covariance. The results were quite similar to the present results (not shown).

The first 10 EOFs of the GS simulations have been used to span the phase space in which the **Had optimal fingerprint** has been computed. This truncation allowed a similarly good representation of both signal guess patterns (on the order of 98%), therefore, augmenting the G only signal pattern into the truncation, as has been done for the MPI truncation space, is not necessary and also not possible, since the χ^2 results indicated unrealistically small model variability if this was attempted (probably due to small-scale noise being the only component of the G simulation not represented by the first 9 EOFs of GS). The χ^2 values started to rise at a truncation level of 20, but were still insignificant. Relatively large differences between the noise properties of Had-ctl1 and Had-ctl2 again suggested an over-emphasis of noise which was merely poorly sampled in the first part of the simulation, and therefore suboptimal results at a truncation level of 12 to 15 (sensitivity tests, again, suggested that the results would have been similar).

To apply optimal GS fingerprints from both model simulations synchronously (Fig. 9), 9 EOFs from HadCM2's GS augmented by the GS signal pattern from HAM3L have been applied. The augmentation with HAM3L GS was necessary to yield a comparable representation of the patterns, and was not rejected by the χ^2 test.

To compare the suppression of noisy signal components between the models, we first analyze how the base space in each model (namely the GS simulation's EOFs) project onto common EOFs (Fig. 3). The first common EOF correlates very highly with HadCM2's second GS EOF ($r = r^* = -0.81$) and HAM3L's second GS EOF ($r = r^* = -0.86$). Therefore, the first common EOF appears to have a counterpart in both model's internal variability. The second common EOF correlates to some extent with the first EOF from the HadCM2 GS simulation ($r = 0.75$, $r^* = 0.50$, this EOF is, however, strongly influenced by the forcing) and mainly through mean warming with the MPI first EOF ($r = 0.64$, $r^* = 0.14$). Higher EOFs cannot be traced in the base space, which is not surprising given that the orthogonality constraint will overwhelm similarities after few EOFs.

We have also compared the weight of each model's EOFs between a simple truncation of the GS signal to that in the optimal fingerprint. We find that each model's first GS EOF (which relates

to the second common EOF) is more suppressed relative to the raw truncation in the Had G fingerprint, while it remains nearly unaltered in the MPI fingerprints. This suggests suppression of tropical signals by the Had fingerprint, which is supported by visual inspection of the optimal fingerprint pattern transformed back to gridbox space (not shown), which shows negative weights over parts of the tropical Pacific. Differently, nearly the entire oceans show positive weights in the HAM3L optimal G fingerprint. The signal components in the second EOF (common EOF 1, which shows the strongest noise in both models) are weighted down by both model's optimal fingerprints, indicating that both model's optimal fingerprints suppress variability in the NH midlatitides. The MPI orthogonal sulfate signal Sul has nearly no weight in EOF 2, while that in HadCM2 has quite strong weight of EOF 2. This corresponds to the stronger similarity of the Had Sul pattern with common EOF 2 than the MPI Sul pattern which could not be fully suppressed by using optimal weights and hinders a detection of the Sul pattern based on the Had fingerprint.

References

Allen MR, Smith LA (1996) Detecting irregular oscillations in the presence of colored noise. J Clim 9: 3373–3404

Allen MR, Tett SFB (1999) Checking for model consistency in optimal fingerprinting. Clim Dyn 15: 419–434

Allen MR, Stott PA (2000) Interpreting the signal of anthropogenic climate change. I: Estimation theory. Tech. Report RAL-TR-99xx, Rutherford Appleton: lab, Chilton, XIIOQX, UK

Allen MR, Stott PA, Mitchell JFB, Schnur R, Delworth LT (2000) Uncertainty levels in forecasts of arthopogenic climate change. Nature, accepted

Barnett TP, Santer BD, Jones PD, Bradley RS, Briffa KR (1996) Estimates of low frequency natural variability in near-surface air temperature. The Holocene 6: 255–263

Barnett TP, Hegerl GC, Santer BD, Taylor K (1998) The potential effect of GCM uncertainties and internal atmospheric variability on anthropogenic signal detection. J Clim 11: 650–675

Barnett TP, Hasselmann K, Chelliah M, Delworth T Hegerl GC, Jones PD, Rasmusson E, Roeckner E, Ropelewski C, Santer BD and Tett SFB (1999) Detection and attribution of recent climate change: a status report. Bull Am Meteorol Soc 80: 2631–2659

Bell TL (1986) Theory of optimal weighting of data to detect climatic change. J Atmos Sci 43: 1694–1710

Boucher O, Lohmann U (1995) The sulfate-CCN-cloud albedo effect. A sensitivity study with two general circulation models. Tellus 47B: 281–300

Collins M, Tett SFB, Cooper C (2000) The internal climate variability of a version of the Hadley Centre coupled model without flux adjustments. Clim Dyn. prov. accepted

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, Hegerl GC, Waszkewitz J (1996) Prediction, detection and regional assessment of anthropogenic climate change. Geophysica 32: 77–96

Cubasch U, Hegerl GC, Voss R, Waszkewitz J, Crowley TJ (1997) Simulation of the influence of solar radiation variations on the global climate with an ocean-atmosphere general circulation model. Clim Dyn 13: 757–767

Delworth T, Knutson K (2000) Simulation of the early twentieth century global warming. Science, 287, 2246–2250

Gordon G, Cooper C, Senior C, Banks H, Gregory J, Johns T, Mitchell JFB, Wood R (2000) The simulation of SST, sea ice extent and ocean heat transport in a version of the Hadley Centre coupled model without flux adjustments. Clim Dyn 16, 147–168

Hasselmann K (1979) On the signal-to-noise problem in atmospheric response studies. In: Shaw BD (ed), Meteorology over the tropical oceans. Royal Meteorological Society, Bracknell, Berkshire, UK, pp 251–259

Hasselmann K (1993) Optimal fingerprints for the detection of time dependent climate change. J Clim 6: 1957–1971

Hasselmann K (1997) Multi-pattern fingerprint method for detection and attribution of climate change. Clim Dyn 13: 601–612

Haywood JM, Ramaswamy V, Donner LJ (1997) A limited-area-model case study of the effects of sub-grid scale variations in relative humidity and clouds upon the direct radiative forcing of sulfate aerosol. Geophys Res Lett 24: 143–146

Hegerl GC, North GR (1997) Statistically optimal methods for detecting anthropogenic climate change. J Clim 10: 1125–1133

Hegerl GC, von Storch, Hasselmann K, Santer BD, Cubasch U, Jones PD (1996) Detecting greenhouse gas induced climate change with an optimal fingerprint method. J Clim 9: 2281–2306

Hegerl GC, Hasselmann K, Cubasch U, Mitchell JFB, Roeckner E, Voss R, Waszkewitz J (1997) Multi-fingerprint detection and attribution of greenhouse gas- and aerosol-forced climate change. Clim Dyn 13: 613–634

Hegerl GC, Jones PD, Barnett TP (2000) Effect of observational sampling error on the detection of anthropogenic climate change. Submitted to J Clim. in press

Hoyt DV, Schatten KH (1993) A discussion of plausible solar irradiance variations, 1700–1992. J Geophys Res, 98: 18 895–18 906

IPPC (1996) Climate change 1995. The IPCC second scientific assessment. In: Houghton JT, Meira Filho LG, Callander BA, Harris N, Kattenberg A, Maskell K (eds) Cambridge University Press, Cambridge, 572 pp

Johns TC, Carnell RE, Crossley JF, Gregory JM, Mitchell JFB, Senior CA, Tett SFB, Wood RA (1997) The second Hadley Centre Coupled Model: description, spinup and validation. Clim Dyn 13: 103–134

Jones A, Roberts DL, Slingo A (1994) A climate model study of the indirect radiative forcing by anthropogenic aerosols. Nature 370: 450–453

Jones PD (1994a) Recent warming in global temperature series. Geophys Res Lett 21: 1149–1152

Jones PD (1994b) Hemispheric surface air temperature variations: a reananlysis and an update to 1993. J Clim 7: 1794–1802

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, Osborn T, Briffa KR (1997) Estimating sampling errors in large-scale temperature averages. J Clim 10: 2548–2568

Jones PD, Briffa KR, Barnett, TP, Tett SFB (1998) High-resolution paleoclimatic records for the last millenium: interpretation, integration and comparison with General Circulation Model control run temperatures. The Holocene 8: 455–471

Jones PD, New M, Parker DE, Martin S, Rigor IG (1999) Surface air temperature and its changes over the past 150 years. Rev Geophys 37: 173–199

Kattenberg A, Giorgi F, Grassl H, Meehl GA, Mitchell JFB, Stouffer R, Tokioka T, Weaver AJ, Wigley TML (1996) Climate models – projections of future climate. In: Houghton et al. (eds) The IPCC second scientific assessment. Cambridge University Press, Cambridge, pp 285–357

Langner J, Rodhe H (1991) A global three-dimensional model of the tropospheric sulfur cycle. J Atmos Chem 13: 225–263

Lean J, Beer J, Bradley R (1995) Reconstruction of solar irradiance since 1600: implications for climate change. Geophys Res Lett 22: 3195–3198

Leroy S (1998) Detecting climate signals: some bayesian aspects. J Clim 11: 640–651

Mann ME, Bradley RS, Hughes MK (1998) Global scale temperature patterns and climate forcing over the past six centruies. Nature, 392: 779–787

Mitchell JFB, Johns TJ, Gregory JM, Tett SFB (1995a) Transient climate response to increasing sulphate aerosols and greenhouse gases. Nature 376: 501–504

Mitchell JFB, Davis RA, Ingram WJ, Senior CA (1995b) On surface temperature, greenhouse gases and aerosols, models and observations. J Clim 10: 2364–3286

Mitchell JFB, Johns TJ (1997) On the modification of greenhouse warming by sulphate aerosols. J Clim 10: 245–267

North GR, Stevens M (1998) Detecting climate signals in the surface temperature record. J Clim 11: 563–577

North GR, Kim K-Y, Shen SSP, Hardin JW (1995) Detection of forced climate signals, Part I: filter theory. J Clim 8: 401–408

Parker DE, Jones PD, Folland CK, Bevan A (1994) Interdecadal changes of surface temperature since the late nineteenth century. J Geophys Res 99: 14 373–14 399

Penner JE, Charlson RJ, Hales JM, Laulainen NS, Leifer R, Novakov T, Ogred J, Radke LF, Schwartz SE, Travis L (1995) Quantifying and minimizing the uncertainty of climate forcing by anthropogenic aerosols. Bull Am Meteorol Soc 75: 375–400

Räisänen J (1997) Climatic response to increasing CO_2 and anthropogenic sulphate aerosols – comparison between two models. Report 46, Department of Meteorology, University of Helsinki, PO Box 4 (Yliopistonkatu 3), FIN-00014, Finland

Räisänen J (1999) Internal variability as a cause of qualitative intermodel disagreement on anthropogenic climate changes. Theor Appl Clim 64: 1–13

Roeckner E, Bengtsson L, Feichter J, Lelieveld J, Rodhe H (1999) Transient climate change simulations with a coupled atmosphere-ocean GCM including the tropospheric sulfur cycle. J Clim, 12, 3004–3032

Santer BD, Brüggemann W, Cubasch U, 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, Penner JE, Wigley TML, Cubasch U, Jones PD (1995) Towards the detection and attribution of an anthropogenic effect on climate. Clim Dyn 12: 77–100

Santer BD, Wigley TML, Barnett TP, Anyamba E (1996a) Detection of climate change and attribution of causes. In: Houghton JT, et al. (eds) Climate change 1995. The IPCC second scientific assessment, Cambridge University Press, Cambridge, pp 407–444

Santer BD, Taylor KE, Wigley TML, Jones PD, Karoly DJ, Mitchell JFB, Oort AH, Penner JE, Ramaswamy V, Schwarzkopf MD, Stouffer RS, Tett SFB (1996b) A search for human influences on the thermal structure in the atmosphere. Nature 382: 39–45

Santer BD, Taylor KE, Wigley TML, Jones PD, Karoly DJ, Mitchell JFB, Oort AH, Penner JE, Ramaswamy V, Schwarzkopf MD, Stouffer RS, Tett SFB, Boyle JS, Parker DE (1996c) Human effect on global climate? Nature 384: 525

Sato M, Hansen JE, McCormick MP, Pollack J (1993) Stratospheric aerosol optical depths (1850–1990). J Geophys Res 98, D12: 22 987–22 994

Senior CA (1999) On the comparison of cloud-climate feedbacks in GCMs. J Clim 12: 1480–1489

von Storch J, Kharim V, Cubasch U, Hegerl GC, Schriever D, von Storch H, Zorita E (1997) A 1260-year control integration with the coupled ECHAM1/LSG general circulation model. J Clim 10: 1525–1543

von Storch H, Zwiers FW (1999) Statistical analysis in climate research. Cambridge University Press, Cambridge, UK

Stott PA, Tett SFB (1998) Scale-dependent detection of climate change. J Clim 11: 3282–3294

Stott PA, Tett SFB, Jones GS, Allen MR, Ingram WJ, Mitchell JFB (2000) Attribution of twentieth century temperature change to natural and anthropogenic causes. Clim Dyn. in press

Stouffer RJ, Hegerl GC, Tett SFB (2000) A comparison of surface air temperature variability in three 1000-year coupled ocean-atmosphere model integrations. J Clim 13: 513–547

Tett SFB, Mitchell JFB, Parker DE, Allen MR (1996) Human influence on the atmospheric vertical temperature structure: detection and observations. Science 274: 1170–1173

Tett SFB, Johns TC, Mitchell JFB (1997) Global and regional variability in a coupled AOGCM. Clim Dyn 13: 303–323

Tett SFB, Stott PA, Allen MR, Ingram WJ, Mitchell JFB (1999) Causes of twentieth century temperature change near the earth's surface. Nature 399: 569–572

Voss R, Sausen R, Cubasch U (1998) Periodically synchronously coupled integrations with the atmosphere-ocean general circulation model ECHAM3/LSG. Clim Dyn 14: 249–266

Wigley TML, Jaumann PJ, Santer BD, Taylor KE (1998) Relative detectability of greenhouse gas and aerosol climate change signals. Clim Dyn 14: 781–790

Willson RC (1997) Total solar irradiance trend during solar cycles 21 and 22. Science 277: 1963–1965

Zhang Y, Wallace JM, Battisti DS (1997) Enso like interdecadal variability: 1903–93. J Clim 10: 1004–1020