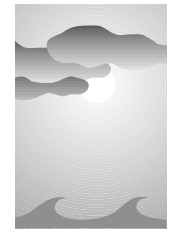
Detection and Attribution of Recent Climate Change: A Status Report



T. P. Barnett,* K. Hasselmann,+ M. Chelliah,# T. Delworth,@ G. Hegerl,& P. Jones,**
E. Rasmusson,++ E. Roeckner,+ C. Ropelewski,## B. Santer,@@ and S. Tett&&

ABSTRACT

This paper addresses the question of where we now stand with respect to detection and attribution of an anthropogenic climate signal. Our ability to estimate natural climate variability, against which claims of anthropogenic signal detection must be made, is reviewed. The current situation suggests control runs of global climate models may give the best estimates of natural variability on a global basis, estimates that appear to be accurate to within a factor of 2 or 3 at multidecadal timescales used in detection work.

Present uncertainties in both observations and model-simulated anthropogenic signals in near-surface air temperature are estimated. The uncertainty in model simulated signals is, in places, as large as the signal to be detected. Two different, but complementary, approaches to detection and attribution are discussed in the context of these uncertainties.

Applying one of the detection strategies, it is found that the change in near-surface, June through August air temperature field over the last 50 years is generally different at a significance level of 5% from that expected from model-based estimates of natural variability. Greenhouse gases alone cannot explain the observed change. Two of four climate models forced by greenhouse gases and direct sulfate aerosols produce results consistent with the current climate change observations, while the consistency of the other two depends on which model's anthropogenic fingerprints are used. A recent integration with additional anthropogenic forcings (the indirect effects of sulfate aerosols and tropospheric ozone) and more complete tropospheric chemistry produced results whose signal amplitude and pattern were consistent with current observations, provided the model's fingerprint is used and detection carried out over only the last 30 years of annually averaged data. This single integration currently cannot be corroborated and provides no opportunity to estimate the uncertainties inherent in the results, uncertainties that are thought to be large and poorly known. These results illustrate the current large uncertainty in the magnitude and spatial pattern of the direct and indirect sulfate forcing and climate response. They also show detection statements depend on model-specific fingerprints, time period, and seasonal character of the signal, dependencies that have not been well explored.

Most, but not all, results suggest that recent changes in global climate inferred from surface air temperature are likely not due solely to natural causes. At present it is not possible to make a very confident statement about the relative contributions of specific natural and anthropogenic forcings to observed climate change. One of the main reasons is that fully realistic simulations of climate change due to the combined effects of all anthropogenic and natural forcings mechanisms have yet to be computed. A list of recommendations for reducing some of the uncertainties that currently hamper detection and attribution studies is presented.

*Scripps Institution of Oceanography, La Jolla, California.
+Max Planck Institute for Meteorology, Hamburg, Germany.
#National Centers for Environmental Prediction, Washington, D.C.
@Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey.
&JISAO, University of Washington, Seattle, Washington.
**University of East Anglia, Norwich, United Kingdom.
++University of Maryland, College Park, Maryland.
##International Research Institute, Lamont-Doherty Earth Observatory, Palisades, New York.
@@Lawrence Livermore National Laboratory, Livermore, California.
&&Hadley Centre for Climate Prediction and Research, U.K. Meteorological Office, Bracknell, United Kingdom.
Corresponding author address: Dr. Tim P. Barnett, Climate Research Division, Scripps Institution of Oceanography, Dept. 0224, 9500 Gilman Drive, La Jolla, CA 92093-0224.
E-mail: tbarnett@ucsd.edu
In final form 13 July 1999.
©1999 American Meteorological Society

1. Introduction

a. Background

The recent meetings in Kyoto, Japan, and Buenos Aires, Argentina, gave a clear signal that many of the world's governments are taking seriously the possibility of substantial changes in planetary climate due to human activities. Yet there has been to date no completely convincing demonstration that the anthropogenic effects predicted by advanced climate models have been unambiguously detected in observations. There are a number of credible studies suggesting such a signal has been detected. But all such statements have been accompanied by major assumptions and substantial caveats. And only recently are detection studies beginning to take account of the inherent model/data uncertainties. The cautious statement by the Intergovernmental Panel on Climate Change (IPCC-95; Houghton et al. 1996) that "the balance of evidence suggests a discernible human influence on climate" admirably summarizes the present ambivalent scientific consensus.

The burden is clearly on the scientific community to demonstrate that their computer scenarios for future climate are realistic. A convincing way to do this is through early detection in the observations of an anthropogenic climate change signal predicted by these state-of-the-art climate models. Until such detection has been convincingly accomplished and the source of some part of the climate change signal clearly attributed to anthropogenic sources, there will be some who simply either do not take the possibility of anthropogenically induced climate changes seriously or who will use it as an excuse for considering remedial action. Further, the detection and attribution statements need to be supported by the majority of the scientific community. It is toward the twin goals of detection and attribution that the current report is directed.

b. Overview of detection and attribution methods

Detecting climate change and then attributing it to specific physical mechanisms is basically a statistical problem, one that can be thought of in terms of familiar regression analysis. Climate models are forced by "scenarios" of how anthropogenic gases have changed in the past and are projected to change in the future. In conventional detection schemes, one takes the time-dependent patterns of spatial change in, say, near-surface air temperature (SAT) from these model runs as the "anthropogenic signal." The observations of SAT are searched for the model-predicted spatial pattern of change. The current detection methods, both conventional and Bayesian (section 4), require that the model-predicted signal, in spatial pattern and amplitude, and its change with time have acceptable counterparts in the observations before detection and attribution (DA) can be claimed. In this case, "acceptable" means that predicted and observed changes are "consistent" to within their (separately) estimated or assumed uncertainties. Most detection methods in use today are "optimal" in the sense that they define, via a rotational transform, the searched-for signal patterns to be as distinguishable as possible from the patterns of natural climate variability. This can simply be thought of as a prefiltering operation of the data prior to application of what are essentially regression techniques (conventional approach) or a set of subjective prior/posterior assumptions via a Bayesian approach. The contrast between these two approaches to DA is elucidated in section 4a.

c. IPCC 1995 report

A recent summary of work aimed at detecting an anthropogenic signal appeared in chapter 8 of the 1995 IPCC report (Santer et al. 1996a). We will not repeat the full discussion but note only the key results and some of the issues that were raised.

The report stated that many previous analyses of global-mean near-surface temperature had concluded that the changes observed over the last century were unlikely to be due to natural variability alone. It was pointed out, however, that it is very difficult to unravel cause–effect relationships in studies of global-mean changes. This is due to uncertainties in our estimates of internally generated and externally forced natural variability (discussed in section 2) and in our estimates of the forcing and response to anthropogenic factors. Given these uncertainties, many combinations of natural and/or anthropogenic effects could yield similar global-mean changes.

It was noted that detection and attribution studies had made major advances since those discussed in the first IPCC report in 1990 (Wigley and Barnett 1990). The advances were judged to be in three main areas: improvements in model-based estimates of an anthropogenic signal, better understanding of the size and characteristics of natural climate variability, and increasing application of "fingerprint" methods that facilitated attribution. While initial detection studies had focused almost exclusively on global-mean changes, or had searched for "greenhouse gas only" signals,

some of the work reported on in chapter 8 utilized pattern-based information from more realistic scenarios of anthropogenic climate change. These involved simultaneous changes in both greenhouse gases (GHGs) and the direct scattering effects of sulfate aerosol particles (SUL). Some of these studies claimed to have detected the model-predicted patterns of response to GHGs and sulfate aerosol direct effects in observed records of near-surface temperature change (e.g., Santer et al. 1995; Hasselmann et al. 1995; Mitchell et al. 1995; Cubasch et al. 1996; Hegerl et al. 1996, 1997). But each of these studies also noted very large uncertainties associated with their conclusions. A summary list of these uncertainties in the IPCC report suggested that the situation was one where an expected signal appeared to be just starting to emerge from (natural) noise, yet was not sufficiently large to be seen clearly given the uncertainties involved. This conclusion was nevertheless somewhat more positive than the 1990 IPCC's detection chapter (Wigley and Barnett 1990), where it was concluded that no definitive statement could be made regarding identification of a human-induced signal in the observed climate records.

d. Outstanding problems and outline

The unresolved problems that faced the detection community after the IPCC report are briefly summarized below. It is this group of problems that the present authors have set about to address and that constitute the outline for the remainder of the paper.

1) Estimating natural variability

Any attempt at detection requires that the amplitude, as well as the temporal and geographic patterns of naturally occurring climate variability, be known or estimated. Without such information, how can one say that recent, observed changes are caused by human activities? They might well have occurred at an earlier time in the historic record for reasons that had nothing to do with human pollution. So in order to claim detection of an anthropogenic climate impact, the change must be significantly different than any likely to be found in, say, the last 1000 years.

2) Accounting for uncertainties

It is obvious that any observations used to detect an anthropogenic signal will be attended by errors, or uncertainties, as we shall refer to them. It is equally clear that the models used to predict anthropogenic impacts will also contain uncertainties due to flaws in their formulation and uncertainties in the way they are forced. These will affect their predicted signals, whose signature we wish to find in the observations. Both classes of uncertainty have only begun to be considered in the most current detection work. Basically, we need to know if we can *detect* a change in climate, given all the model/data uncertainties. If we can say climate has changed, can we then *attribute* the change to specific forcing mechanisms, given the level of model/data uncertainty?

3) Methodology

The detection methods reported in IPCC-95, chapter 8, essentially sought to find in the observations some type of model signal that represented a spatial pattern of change in, say, near-surface temperature. The temporal characteristics of this spatial pattern, for example, averaging interval, seasonal dependence, etc., were not fully considered. Perhaps more fundamental, the detection issue was couched in a rigid framework that allowed no room for a priori assignment of uncertainties on the signal and/or the data used to detect it. This excluded the use of potentially valuable qualitative climate change indices or events such as glacier melting, sea ice changes, etc.

We have investigated all three of the areas described above. The remainder of the report describes our current findings in these areas. These findings have been incorporated into our detection strategy. This improved strategy has been used to provide an assessment, in the final sections, about our current confidence in statements that an anthropogenic signal has, or has not, been detected. This is largely based on including the quantified uncertainties as fully as possible into one particular detection approach (Hegerl et al. 1997). Our results emphasize that future detection approaches must focus more strongly on and include the uncertainty estimates if they are to be useful.

As an aid to readers, each of the following sections (2–5) contains a "purpose" statement at its beginning and "summary" statement(s) at its end. A superficial overview of the paper can be obtained by reading these parts alone. Sections 6 and 7 summarize the main results and suggested future research, respectively.

2. Estimating natural variability

To attribute an observed climate change to human influence, it is first necessary to show that such a climate change is unlikely to have occurred naturally, for example, has not occurred in the past. For purposes of

this paper, we will confine our attention to climate changes as manifest in the SAT field, since it has long, near-global coverage. There are at least two sources of natural variability in SAT: Internal interactions within the climate system, such as ENSO, and external forcing, for example, through changes in solar luminosity or volcanic activity. In the following we shall refer to both externally forced and internally induced fluctuations as natural climate variability.

So how does one know what has happened before industrialization? There are three general approaches to answering this question. The most obvious approach is to look at the instrumental record of SAT, the climate variable that has been best observed. One can also look at paleoclimate proxies or, alternatively, use climate models to infer "normal" climatic behavior of SAT. Each of these approaches is discussed below.

a. Instrumental records

Direct estimation of the natural variability of SAT from actual observations has a number of problems. With a few notable exceptions, local records are generally less than 100 years in length. As we shall see, detection requires analysis over periods of the order of several decades or more, so the instrumental record has only a few realizations of climate change on these timescales. This is not adequate. Further, the instrumental record must be contaminated by an anthropogenic signal if one is present. Attempts to remove such a signal have been made by Jones and Hegerl (1998), among others (see also Wigley et al. 1998b). Such anthropogenic signal removal or "correction" leaves a residual that can be attributed to natural variability. The results show that the corrections are a sizable fraction of the signal itself, a highly unsatisfactory situation, especially when the correction depends on an assumed model that itself may be uncertain (Fig. 1).

The instrumental data may also have biases due to changing measurement methods/instruments. These biases are often poorly documented. Although they can be as large locally as the currently expected anthropogenic signal, it is believed that they have been adequately corrected for in the SAT dataset (Jones et al. 1999). Finally, there are large expanses of the planet where the observations are either scant or missing altogether, especially during the early part of this century. Many of these areas are just where an anthropogenic signal is expected to be most prominent. The impact these data-void regions have on detection was discussed by Wigley et al. (1998a). Detection and attribution studies remove these data-poor regions from the simulated signals so that only information from data-adequate areas is used.

In summary, while instrumental data are marginally adequate for validating climate models and first detection efforts, they cannot be used for estimating multidecadal or secular natural variability of SAT over the last few centuries or more.

b. Paleoclimate proxies

Tree rings, coral records, and ice cores, among others, have been suggested as proxies of past climates. They certainly have the longer records, extending back in many cases 500–1000 years, required to estimate natural climate variability on multidecadal scales. And almost all of this record is before there was any possible contamination from anthropogenic sources. Indeed, the proxies appear to offer the only means of directly assessing natural variability prior to instrumental records.

Unfortunately, there are numerous problems with using paleodata to estimate past climate changes in the SAT. The proxies are largely indicators of local climate change and are sparsely scattered over the globe. Perhaps more importantly, they are generally not per-

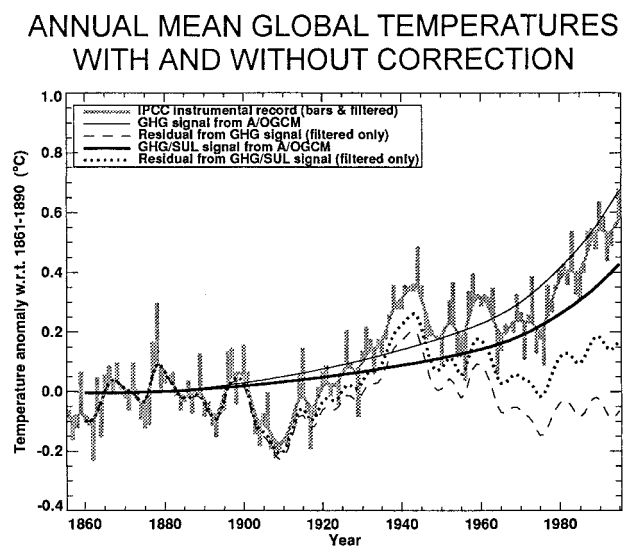


FIG. 1. Annual-mean observed global temperatures (expressed as anomalies from 1861 to 1890) for 1856–1995 (gray histogram and 10-yr Gaussian filter). The thick and thin smooth lines, respectively, are the anthropogenic and greenhouse-gas-only fitted responses (from Jones and Hegerl 1998). The model responses come from Hegerl et al. 1997. The dotted and dashed lines, respectively, are the "residuals" after extracting the anthropogenic and greenhouse-gas-only signals from the 10-yr Gaussian smoothed data.

fect temperature indicators, but rather reflect a blend of different climatic effects. For example, coral records might demonstrate the combined effects of changes in ocean temperature, precipitation and its runoff, sea level changes, and turbidity of the waters in which the corals live. Dating uncertainties and low-frequency distortion introduced in climate reconstructions for trees, corals, and ice cores are also present in paleorecords (Jones et al. 1998).

Recent compilations of paleoclimatic data have offered the first opportunity to analyze this type of data on a global scale. Straightforward comparisons via cross-spectral analysis of the recent paleodata with the instrumental record show that most of the paleodata are not simple proxies of temperature (Barnett et al. 1996; Jones 1998; Jones et al. 1998; see Table 1). Indeed, only a few of the tree-ring records from mid- to high-latitude sites can be interpreted directly as temperature changes. Attempts by Jones et al. (1998) to use these "good" records to construct a record of Northern Hemisphere (NH) temperature over the last five centuries are shown in Fig. 2. Also shown is a different reconstruction created using a full compilation of proxy data (Mann et al. 1998). The disparity between these reconstructions at some times over the last 400 years is as large as the observed changes in global temperature over the last 100 years. Some of the differences are due to different compilations of proxy data and also differences in the seasons reconstructed, but most of the disparity simply represents uncertainties in our knowledge of past changes in NH average temperature.

The general conclusion from the above, and other studies, is that for a few regions and some proxies, reliable reconstructions of temperature change on the required decadal timescales can be produced. The number, however, is quite limited in relation to the volume of paleodata available. At present, it is debatable whether there is enough temperature proxy data to be representative of hemispheric, let alone global, climate changes given the lack of large spatial scale coherence in the data. Yet the few good records that are available serve as strong checks on efforts to model natural climate variability (Jones et al. 1998).

c. Climate models

Global climate models, operated in a control-run mode, may offer the best chance to estimate natural variability. Model simulations can be long and provide uniform global coverage of many different variables, just what is needed to estimate internal climate variability.

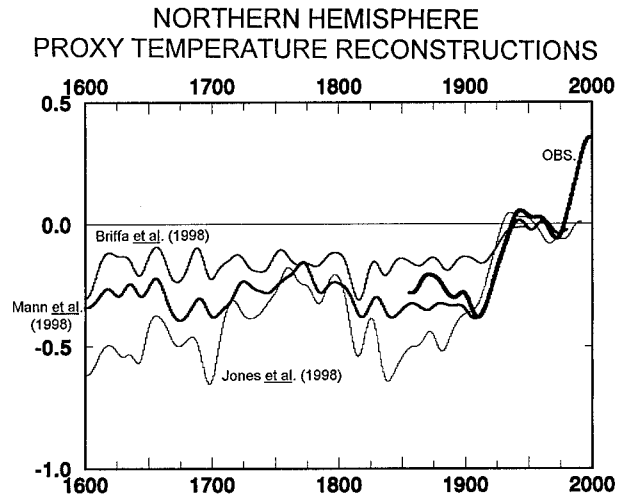


FIG. 2. Comparison of three Northern Hemisphere proxy temperature reconstructions with instrumental data on the same space scale. The thinnest line is from Jones et al. (1998) and is based on a simple average of 10 NH proxy reconstructions. The next thinnest line (generally the highest) comes from a tree-ring density reconstruction in Briffa (1998) using up to 400 series from the high latitudes and high elevations of the NH. The thicker line is from Mann et al.'s (1998) multiproxy assemblage. The thickest of all lines is the annual temperature average for the NH (based on both the land and the marine components) for the period 1856–1998. All lines have been smoothed with a 30-yr Gaussian filter and all are expressed as anomalies from 1961 through 1990.

The major question here is, Can coupled global climate models (CGCMs) accurately reproduce natural low-frequency variability in SAT (or any other variable)? New studies, based on the recent observational record, suggest the answer to this question for SAT is a qualified yes. For instance, power spectra of global mean temperature from four different CGCMs[1] agree moderately well with that of the observations in the frequency range corresponding to the averaging times, 20–50 yr, used in detection studies (Fig. 3, and Stouffer et al. 1999). However, the differences in energy in this key frequency band still varies by a factor of 2–3 to between models. The low number of degrees of freedom and large error bands in the spectral analysis make more exact estimates impossible. Joint eigenanalysis of multiple, long CGCM control simulations, paleoproxies, and observations by Stouffer et al. (1999), Barnett et al. (1996), and Jones et al. (1998) suggest that the models reproduce the observed spatial–temporal structure of near-surface temperature

[1] The GFDL model used here had R15 resolution. The new GFDL model used later in the paper has R30 resolution.

TABLE 1. Cross-spectral analysis of paleoproxies and collocated observed near-surface temperatures.

Data source/ location	Type	Coherence squared: paleo vs. instrumental* Period band (yr) 40	20	10	5	4	3	Correlation
Svalbard	Ice	0.19	0.13	0.14	0.9	0.25	0.05	0.22
S. Greenland	Ice	0.16	0.09	0.10	0.26	0.17	0.24	0.07
Crete	Ice	0.26	0.48	0.50	0.24	0.54	0.43	0.50
Law Dome	Ice	0.19	0.12	0.00	0.25	0.16	0.01	0.13
GB Reef	5 corals	0.46	0.19	0.31	0.26	0.09	0.33	0.18
GB Reef	10 corals	0.46	0.26	0.10	0.07	0.26	0.38	0.27
Galp. Island	Coral	0.05	0.20	0.22	0.32	0.64	0.16	0.32
Kapoposa	Coral	0.52	0.25	0.23	0.22	0.09	0.05	0.38
New Cal.	Coral	0.40	0.40	0.47	0.36	0.31	0.08	–0.20
N. Treeline	Tree	0.24	0.13	0.07	0.05	0.16	0.04	0.15
W. U.S.	Tree/width	0.74	0.66	0.52	0.59	0.14	0.15	0.43
W. U.S.	Tree/den	0.51	0.41	0.77	0.55	0.60	0.66	0.69
New Zealand	Tree	0.17	0.12	0.39	0.27	0.23	0.90	–0.15
N. Fenno.	Tree	0.63	0.45	0.43	0.36	0.56	0.73	0.72
N. Urals	Tree	0.93	0.65	0.67	0.54	0.65	0.43	0.79
Jasper	Tree	0.20	0.12	0.53	0.31	0.04	0.08	0.49
Tasmania	Tree	0.66	0.50	0.13	0.05	0.54	0.20	0.55
Chile	Tree	0.07	0.16	0.03	0.20	0.07	0.50	0.29
Argentina	Tree	0.05	0.26	0.21	0.31	0.52	0.12	0.40
C. England	Instru.	0.73	0.56	0.83	0.86	0.56	0.86	0.88
C. Europe	Doc.	0.83	0.77	0.70	0.54	0.59	0.94	0.89

*Data from Jones et al. (1998), which also contains reference to paleo data sources. Underlined coherence squared values are significant at the 95% level.

changes surprisingly well, although the pattern loadings between models and observations also varied here by about a factor of 2–3.

The major shortcomings found in these studies is a tendency for the models to underestimate the magnitude of the largest spatial scales of variability. This is shown in Fig. 4 where the model data projected onto its common EOF eigenstructure typically have less energetic eigenvalues than those obtained from the observations projected onto the same basis set. This

could be due mainly to the fact that the runs studied had no forcing from either anthropogenic or natural sources. When estimates of solar and volcanic forcing are added into the CGCM control simulations, the underestimation problem is reduced (e.g., Cubasch et al. 1997). We shall see in section 5 that simulations with these CGCMs reproduce the near-surface air temperature changes of the last 100 years rather well.

d. Summary

Claims of detection of an anthropogenic climate change signal, for example, in the SAT field, must compete against the likelihood that the observed change is due to natural causes. It appears that our best estimates of natural variability will come from CGCMs, even if the variance levels can be determined only to within a factor of 2–3. Instrumental and paleoproxy data have serious shortcomings that preclude their use for this purpose. However, these latter data types can provide valuable validation checks on the CGCMs, building up enough confidence in the models to use them for estimates of natural variability in detection studies.

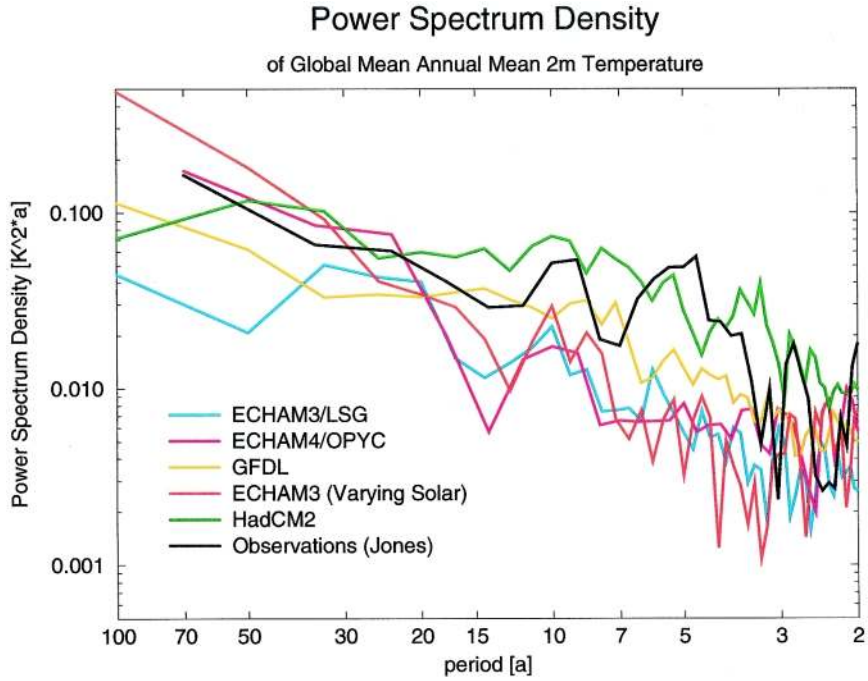


FIG. 3. Power spectra of global-mean, annual temperatures from a number of coupled model control integrations compared with that derived from observations. The model designators are explained in Table 3. The record lengths vary and so do the confidence limits. In most cases, the confidence limits are large compared to differences between spectra. Note the difference in energy in the 30–50-yr band, where most detection work is done, is of order 2–4. See Stouffer et al. 1999 for additional details.

3. Uncertainties

Only recently has detection work paid serious attention to the variety of uncertainties that attend the observations and model projections of an anthropogenic signal. This section briefly discusses and illustrates several of the more significant uncertainties; space limitations preclude a more detailed discussion of this important issue here. The impact of uncertainties on detection statements will be illustrated in section 5 by sensitivity tests with one particular detection and attribution method.

a. Observational uncertainties

1) NEAR-SURFACE TEMPERATURE

Near-surface air temperature has been the major variable used in prior detection studies because it is a fairly easy measurement that has a long historical record (e.g., Barnett et al. 1991; Santer et al. 1995; Hegerl et al. 1996). These data have been used to produce time series of changes in global and hemispheric mean temperatures over the last 100 or so years. It is only recently that the magnitude of the sampling errors associated with these calculations have been included on the time series (Fig. 5, after Jones et al. 1997; see also Karl et al. 1994). Typical 95% confidence limits for estimates of the global means on interannual timescales are 0.11°C since 1951 and 0.17°C prior to that date. The uncertainty values are significantly higher for regions and higher still for individual 5° grid boxes.

It is often stated that the global temperature has been increasing steadily since the turn of the century. Figure 5, with the confidence limits included, shows this is not an accurate statement. Rather, the temperature increased abruptly between about 1920 and 1945 and again from 1975 to the present (e.g., Jones et al. 1999). The confidence limits are large enough so that no change in temperature can be claimed outside these periods, for example, between about 1860 and 1920 and about 1940 and 1975. Nevertheless, the observational uncertainties are considerably smaller than the

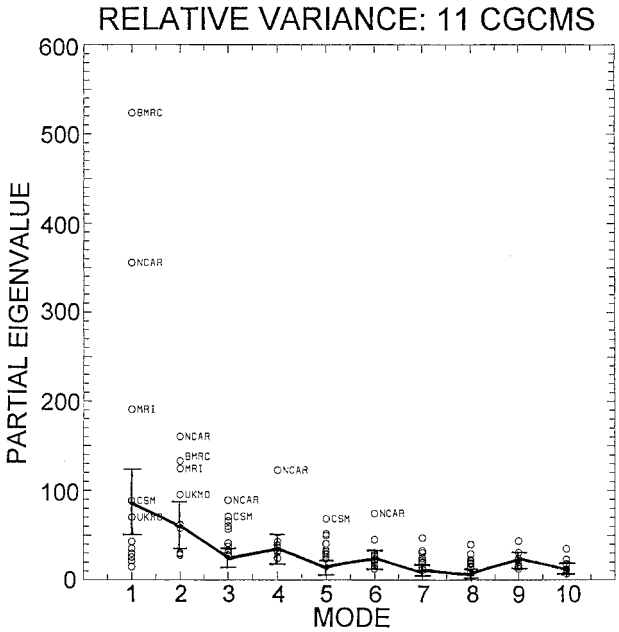


FIG. 4. Partial eigenvalue spectrum. The letter codes refer to various coupled global climate models. The heavy solid line represents the partial eigenvalue spectrum obtained by projecting the observed, detrended air temperature onto the CGCM common EOF basis set. The vertical bars show the approximate 95% confidence limits on the observed partial eigenvalues (after Barnett 1998).

total change since 1900, so the increase in global-mean temperature is highly statistically significant. Although, as pointed out, these data could potentially contain contamination due to urban heat island effects and biases (especially in ocean temperatures) due to changing instrumentation (e.g., Barnett 1984), recent studies suggest these errors have probably been adequately allowed for (Jones et al. 1990; Parker et al. 1995).

2) UPPER-AIR TEMPERATURES

Some of the first studies to claim qualified detection of a model-predicted anthropogenic signal were based on the free air temperatures (Santer et al. 1996b; Tett et al. 1996). These data came from radiosondes and did not really have reasonable spatial coverage until the late 1950s, and even then were highly irregular in time (Gaffen et al. 1999), providing a relatively short record, at best about 40 years, for detection analysis. Changes in instrumentation are known to have introduced very large discontinuities to this dataset (e.g., Parker et al. 1997). However, comparison of lower-tropospheric temperature trends from radiosondes and the satellite-derived Microwave Sounding Unit (MSU) 2R over the period 1979–96 is claimed to be good (e.g., Christy et al. 1998). Unfortunately, the satellite dataset is only 20 years long and therefore too short for practical detection work. In any event, such comparisons typically have neglected large uncertainties in the radiosonde datasets, particularly during the 1958–78 period, and have not accounted for data coverage differences. A more definitive discussion of the upper-air data and its problems can be found in Parker et al. (1997) and Santer et al. (1999). For present purposes, it appears that radiosonde data, despite their deficiencies, have reasonable horizontal and vertical resolution, which eventually may be of considerable help in constraining present uncertainties in anthropogenic forcing and in model-based signal estimates.

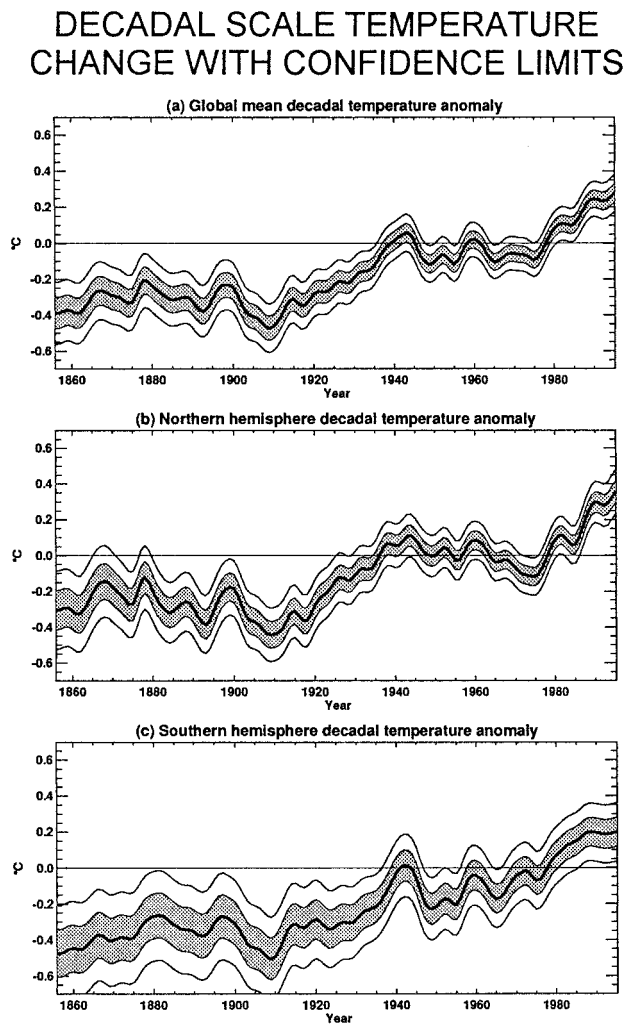


FIG. 5. Decadal timescale surface temperature record for (a) global, (b) Northern Hemisphere, and (c) Southern Hemisphere means, with ±1 standard error (shaded) and ±2 standard error (thin lines). Temperatures expressed as anomalies from the 1961–90 mean. After Jones et al. (1997).

3) Reanalysis products

Extensive efforts have recently been made in the area of reanalysis. Dynamical atmospheric general circulation models have been applied to assimilate available observations in an attempt to produce a temporally homogeneous, regularly gridded, dynamically consistent reconstruction of the 3D atmospheric structure over the last 20–40 years [Kalnay et al. (1996) for the National Centers for Environmental Prediction (NCEP) and Gibson et al. (1997) for the European Centre for Medium-Range Weather Forecasts (ECMWF)]. It is natural to inquire if these reanalyses data can be used for DA.

Unfortunately, our studies indicate that the current generation of reanalyses have very limited usefulness for detection and attribution studies. For example, the NCEP and ECMWF reanalyses show inhomogeneities and very different temperature behavior in the lower and midtroposphere (Santer et al. 1999; Chelliah and Ropelewski 1999, manuscript submitted to *J. Climate,* hereafter CR). In the former region, the temperature trends in the two reanalyses differ by up to 0.11°C/decade over 1979–93—a value within the estimated range of an expected anthropogenic signal (Santer et al. 1996a). NCEP and ECMWF also have large systematic differences in the midtroposphere, particularly in the Tropics (Fig. 6; CR). Also, Santer et al. (1999) have compared layer-average temperatures in NCEP and ECMWF with those in MSU (versions b, c, and d) and various radiosonde datasets and also find serious problems with both sets of reanalysis. In summary, use of reanalysis data in DA studies will have to await a future generation of reanalysis products.

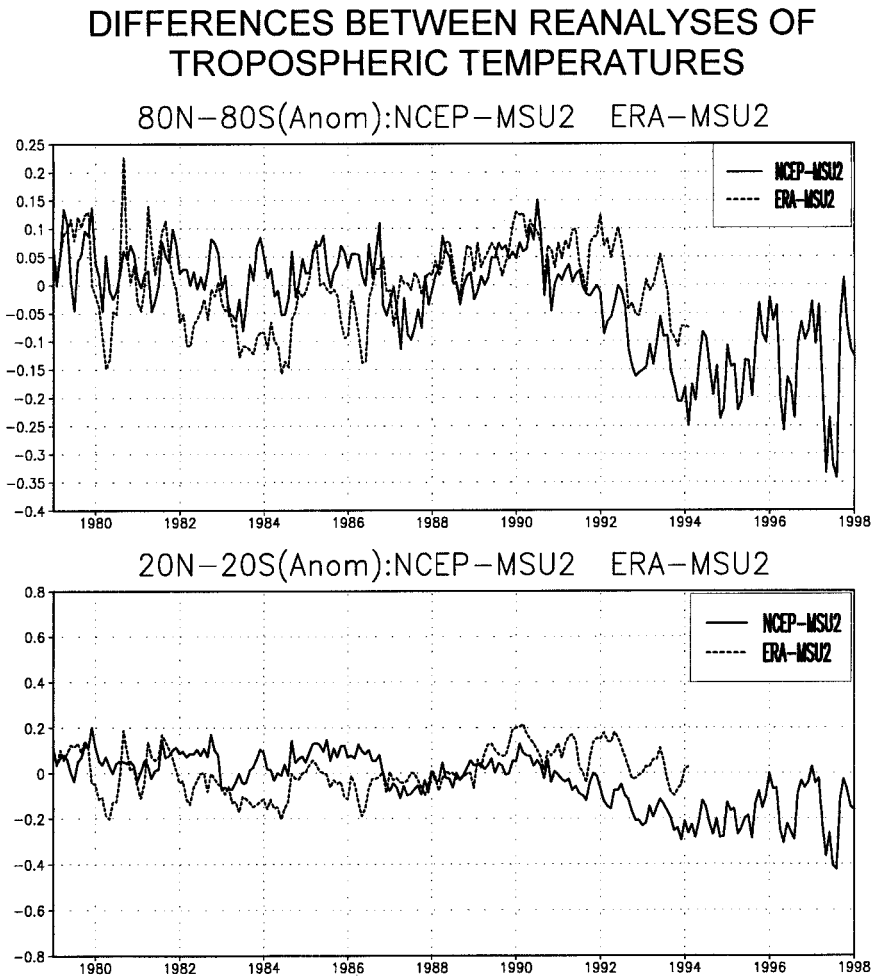


Fig. 6. Time series of the tropospheric temperature anomaly differences between NCEP reanalysis and MSU2 (solid) and ECMWF reanalysis and MSU2 (dotted). (top) The globe equatorward of 80° latitude. (bottom) The Tropics equatorward of 20° latitude. After CR.

b. Model uncertainties

The model-predicted anthropogenic signal is associated with a number of uncertainties. Some are due to fundamental errors in the models themselves, for example, an inability to produce a credible ENSO cycle, parameterizing ocean mixing, clouds, etc. Others arise from errors in the forcings that are included (such as sulfate aerosol direct effects), and from errors due to neglect or inadequate specification of poorly known anthropogenic or natural forcings, such as indirect sulfate aerosol effects and volcanic dust loadings. Another major source of uncertainty arises from the expected internal model variability, for example, nonlinear interactions within the models that produce large variability even when run in control mode (no external forcing of any kind). In the brief space we have here, it is only possible to give examples of several of these uncertainties as they relate to detection of a large-scale anthropogenic signal. In section 5, we will see what impact they have on detection statements.

1) Anthropogenic signal differences

How much uncertainty is introduced into the detection problem by different representations of the same anthropogenic forcing in different computer models? An idea of the difference was obtained by analyzing the ensemble of scenario runs produced by the Hadley Centre (HC), the Geophysical Fluid Dynamics Laboratory (GFDL), and the Max Planck Institute of Meteorology (MPI). These runs simulated the time-dependent effects of increasing "total equivalent" CO_2 (denoted GHG) and the direct effect of sulfate aerosols (denoted SUL), the latter represented in the models by a change in surface albedo. So all the models were forced with conceptually the same anthropogenic pollutants. Just how the forcing from these pollutants was specified varied from model to model. The details of the models, ensembles and appropriate references are given in Table 2.

The (smoothed) temperature changes between 1945 and 1995 predicted by the ensemble average for each model simulation are shown in Fig. 7. All of the models suggest warming over most of the oceans and continents of order 0.5°C or less. The main difference between models (Fig. 8) is over the midlatitude oceans of the Northern Hemisphere and over or just east of the main industrial sectors of European and North American landmasses. The main differences, which can easily exceed 0.5°C locally in the models' predicted anthropogenic signal, could be due to the way in which models represent sulfate aerosols and other physical differences such as clouds (e.g., Hegerl et al. 1999). Fortunately, most of the disparity between the ensemble average of the models is high wavenumber in character, while current detection schemes use only the very low wavenumber information, for example, Stott and Tett (1998). In this case, the dissimilarities between models will appear as noise and might not greatly affect standard detection methods. On the other hand, Fig. 8 does have some large-scale features and these could impact DA studies (e.g., lower panel).

2) Internal model variability

Each CGCM demonstrates relatively large variability generated by internal nonlinear interactions. This is demonstrated in Table 3 where the pattern correlations between various of the realizations listed in Table 2 show large intramodel variability between predicted June–July–August (JJA) trend patterns over the period 1945–95. Estimates of this variability suggest it is comparable with the observed natural climate variability (Barnett 1999; Manabe and Stouffer 1997; Barnett 1995; Stouffer et al. 1999; among others).

Table 2. Model descriptions.

ID	Model name	Modeling center	Grid size (°)	CTL run (yr)	Scenario runs (number and type)
HadCM2	HadCM2	Hadley Centre, Bracknell, UK	A: 3.75 × 2.5 O: 3.75 × 2.5	~1610	Four G, four GS, four Sol, four Vol
HAM3L	ECHAM3/LSG	MPI, Hamburg, Germany	A: 5.625 × 5.625 O: 3.5 × 3.5	~1518	G, two GS, two Sol
HAM4P	ECHAM4/OPYC	MPI	A: 2.8 × 2.8 O: variable	~300	G, GS, GSI
GFDL	GFDL R30	GFDL, Princeton, NJ	A: 3.75 × 2.25 O: 1.9 × 2.25	~675	Five GS

G = greenhouse gas only.
GS = G + direct sulfate aerosol forcing.
GSI = GS + indirect aerosol and troposhere ozone forcing.
Sol = time-dependent solar insolation forcing after Hoyt and Schatten (1993).
Vol = time-dependent volanic forcing after Sato et al. (1993).
The grids of the ocean (O) and atmospheric (A) models are given. Details of the HadCM2 model are given in Johns et al. (1997), the MPI model in Voss et al. (1997), and the GFDL model in KDDS.

Earlier work by Cubasch et al. (1994) demonstrated this problem could be important particularly for early detection of an anthropogenic signal. In their simulations, the early evolution of the anthropogenically induced signals differed markedly for integrations starting from different initial states.

We illustrated this using the four HC (model HadCM2) simulations described in Table 2. Each used the same model and the same representation of anthropogenic forcing, the only difference between the runs being perturbations in the initial conditions. Hence, the differences between the runs were due solely to internal model variability. The matter of concern is the size of the model's anthropogenic signal compared to the differences between individual signal estimates. To determine this ratio, the individual runs were first smoothed in time to eliminate "end point" effects and the temperature change between 1995 and 1945 was computed. The mean signal was determined by averaging together the four realizations of smoothed temperature change (Fig. 9, lower panel). At each grid point, we then computed the ratio of the mean signal to the standard deviation between the four runs (Fig. 9, upper panel). The regions in Fig. 9 (upper panel) where the ratio is greater than one are colored. In these areas the signal exceeds the "noise" associated with internal model variability, while in the remaining gray regions the ratio is one or less. In these areas, the mean model GHG+SUL signal based on only four realizations is uncertain.

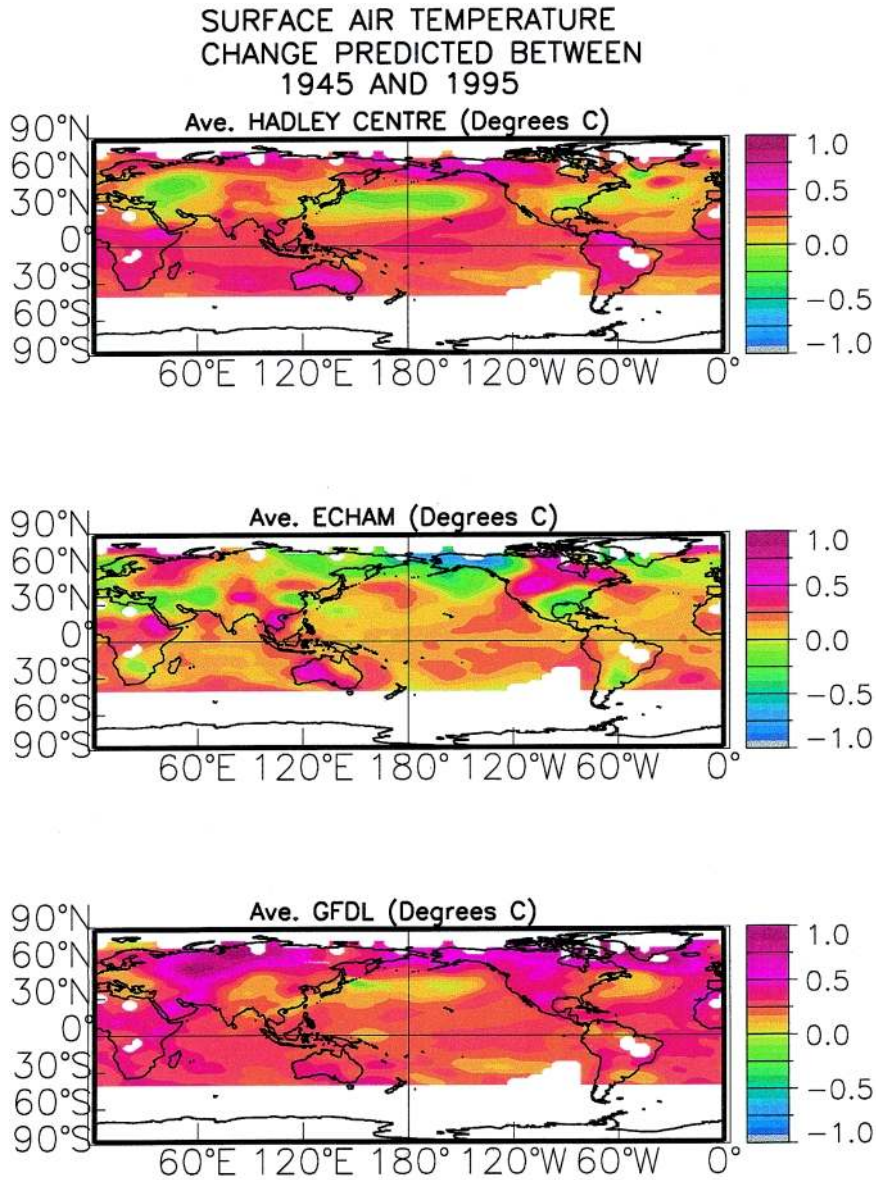


FIG. 7. Changes in smoothed near-surface temperature (°C) between 1945 and 1995 produced by ensemble averages from coupled global climate models of the HadCM2, MPI, and GFDL. All models were forced by their own independent estimates of an anthropogenic signal composed of greenhouse gases and direct sulfate aerosol effects.

The results show that, for the 1945–95 period, perhaps half the Northern Hemisphere above 30° latitude has a mean anthropogenic signal that is equal to or less than the intramodel noise. This applies particularly to the aerosol fallout regions noted above. Detection of the signal in these regions at this time will be difficult. In contrast, strong, clear anthropogenic signals are observed over the continents, especially those in the Southern Hemisphere, and all of the Tropics.

The results discussed above, plus those given by Knutson et al. (1999, manuscript submitted to *J. Geophys. Res.,* hereafter KDDS) and Barnett et al. (1999, manuscript submitted to *J. Geophys. Res.,* hereafter BHKT) illustrate the dangers of using only a single CGCM anthropogenic run in early detection studies. In its early stages of growth, the space–time

structure of the signal can be distorted by the internal model variability[2] thereby most likely depressing or elevating the detection significance levels. It is current practice to use ensembles of scenario runs in DA studies to minimize this possibility. Unfortunately, this cannot be done with the (noisy) observations of which we have only one realization.

[2] The most recent studies explicitly include estimates of these uncertainties as will be demonstrated in section 5.

c. Summary

Uncertainties that are large relative to the *currently* expected anthropogenic signal exist in the observations. In most cases, these uncertainties cannot easily be eliminated since they are associated with instrumental changes over the century. The sampling uncertainties, however, can be estimated and are relatively small for the SAT observations currently used in most detection studies. Similar uncertainties arise through the models used to estimate the anthropogenic signal. These uncertainties are also poorly known, since they have been investigated in only a few models, and even then not completely.

Thus, to avoid overoptimistic (or overpessimistic) claims of detection, especially early detection, it is essential that a careful assessment of all potential error sources in both observations and models be made and their impact properly included in the detection methodology. The inherent uncertainties involved in making these error estimates must be emphasized. These factors are the origin of many of the caveats permeating the published literature on detection.

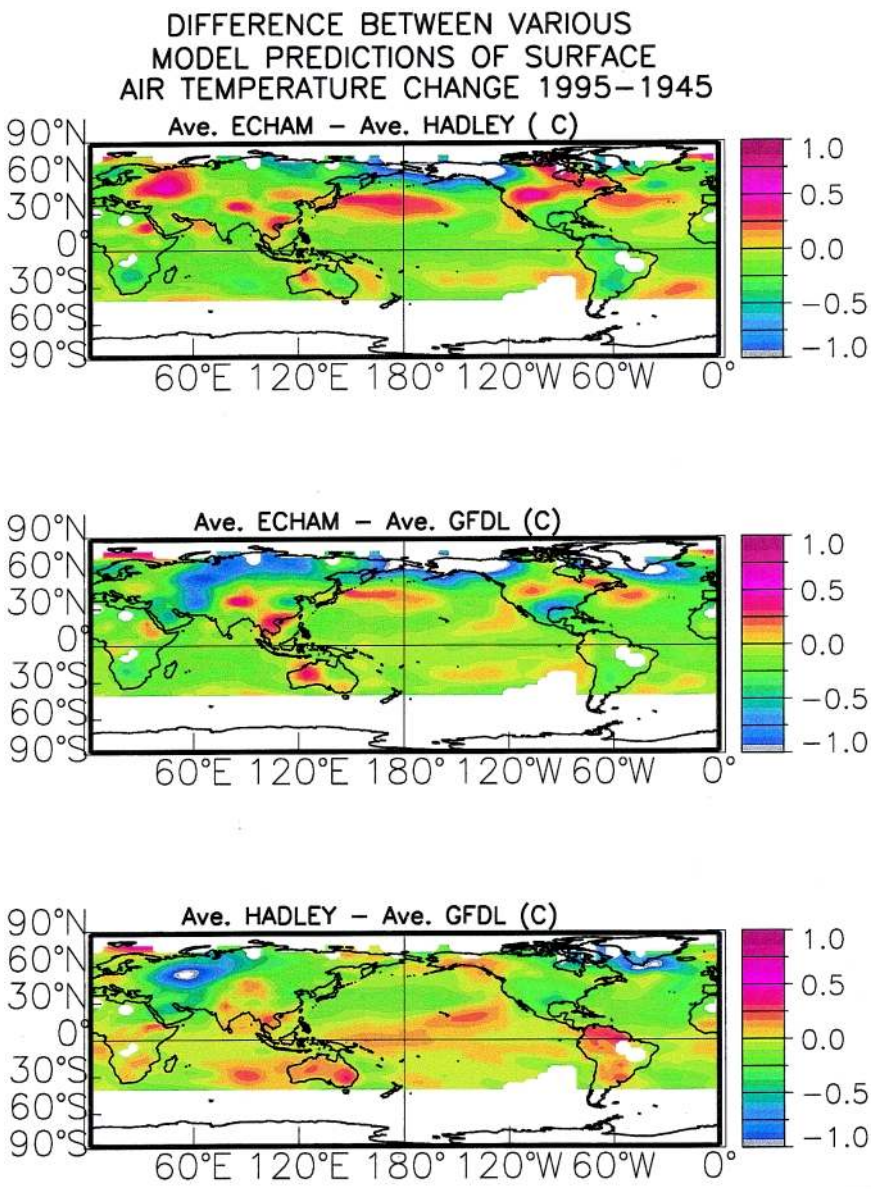


FIG. 8. Differences in ensemble averaged temperature change (°C) between the three model scenarios described in Fig. 7. Note the local values of the difference fields are often as large or larger than the mean change in any one model (Fig. 7). However, the spatial scale of the difference is somewhat smaller than those associated with the mean field.

4. Theory

We summarize in the following the basic concepts and methods of application of the conventional and Bayesian approaches to detection and attribution. Both approaches are based on formal statistical theory, but differ in how they translate the results of the formal theory into the real world. Details of the general mathematical formalism of both approaches are given in appendixes A and B. Readers more interested in the results of applying these theories may wish to skip to section 5 describing the current status of DA studies.

TABLE 3. Pattern correlation of spatially smoothed JJA temperature trends between various realizations of GHG+SUL forcing from different models. Trend Period: 1945–95. See Table 2 for model references.

	HC 1	HC 2	HC 3	HC 4	EC A	EC B	GF 1	GF 2	GF 3	GF 4	GF 5	Obs.
HC 1	1.00	0.59	0.28	-0.08	0.42	0.50	0.52	0.38	0.26	0.23	0.36	0.26
HC 2	0.	1.00	0.33	0.21	0.37	0.38	0.59	0.14	0.41	0.46	0.49	0.21
HC 3	0.	0.	1.00	-0.14	0.67	0.51	0.40	0.50	0.04	0.58	0.14	0.25
HC 4	0.	0.	0.	1.00	-0.33	-0.16	-0.08	-0.22	-0.16	0.03	0.07	-0.29
EC A	0.	0.	0.	0.	1.00	0.68	0.48	0.45	0.33	0.55	0.26	0.23
EC B	0.	0.	0.	0.	0.	1.00	0.43	0.54	0.13	0.29	0.08	0.22
GF 1	0.	0.	0.	0.	0.	0.	1.00	0.41	0.32	0.56	0.37	0.04
GF 2	0.	0.	0.	0.	0.	0.	0.	1.00	0.14	0.25	0.00	0.17
GF 3	0.	0.	0.	0.	0.	0	0.	0.	1.00	0.13	0.50	0.07
GF 4	0.	0.	0.	0.	0.	0.	0.	0.	0.	1.00	0.13	0.02
GF 5	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	1.00	0.16
Obs.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	1.00

HC= Hadley Centre/HadCM2.
EC= MPI/HAM3L.
GF= GFDL/R30.
Obs.= observations.

a. Conventional and Bayesian statistics: Overview

Formal statistical theory is concerned with probability distributions of variables or fields defined with respect to an infinite ensemble of realizations. In practice, we never have such an infinite ensemble of realizations, that is, the true frequency distribution of a variable is not known. This is especially true in detection work where this theoretical requirement is satisfied only marginally for the natural variability of near-surface temperatures. However, we normally assume that the observed temperature distributions approximate the unknown abstract distribution. In this case, if the CGCM-predicted near-surface temperature signal is known, signal-to-noise ratios can be computed using conventional statistics. On the other hand, the impact of the signal uncertainty associated with model errors currently lies outside the scope of conventional statistical theory, as there exists only the small set of model ensembles described in section 3b on which to base the model error statistics. Similarly, conventional statistics cannot be applied to variables for which there exist no reliable estimates of the natural variability on the range of timescales relevant for detection, say, the waxing and waning of glaciers over the last 500 years. This is a severe restriction on the conventional method, since there exist many indicators of climate change that suffer from this limitation, but that one would nevertheless like to incorporate in a comprehensive climate change detection and attribution study.

These shortcomings are overcome in the Bayesian approach, albeit at the expense of a subjective rather than an objective definition of probability of detection. The Bayesian probability concept is based on the ev-

eryday observation that people speak of probabilities (and are willing to quantify their probability assessments through investments in the stock market, e.g.) independent of the existence of frequency distributions. It is assumed that people have a rational basis to assign to events or hypotheses, such as the existence of an anthropogenic climate change signal, subjective probabilities based simply on past experience and limited available information. New evidence, such as the analysis of an observational dataset, results in the conversion of a person's prior probability assessment into a posterior probability. In addition to opening the door to climate change indices that would otherwise be excluded from a detection study based on the conventional statistical approach, the Bayesian approach offers a rational framework for the debate on controversial issues of detection and attribution that are often strongly colored by individual, subjective estimates of model errors.

b. The optimal fingerprint method of conventional statistics

In the conventional approach, the analysis is carried out in what can be called "attribution space," a low-dimensional subspace derived from the model-predicted signals. Each dimension of this space can simplistically be thought of as associated with the fingerprint of a particular type of anthropogenic forcing, for example, greenhouse gases, sulfate aerosols, etc. However, the fingerprints are not identical to the signal patterns, but are obtained from these by a rotation in the climate phase space away from the directions with the highest natural variability noise. This is accomplished by a simple inverse weighting of the observations/signal with the covariance of natural variability (estimated from long CGCM control runs). The search for the signal is then carried out in the optimal fingerprint direction that maximizes the signal-to-noise ratio (see appendix A). The rotation can be viewed as a filtering operation analogous to the standard practice of prewhitening time series data.

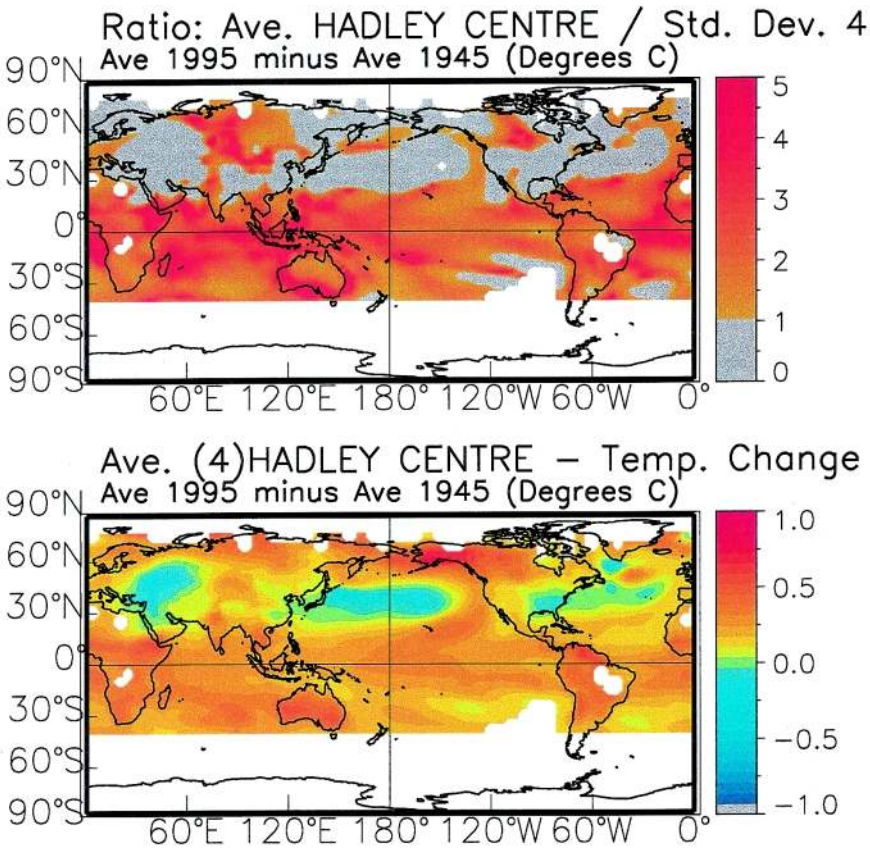


FIG. 9. (bottom) Ensemble average temperature change 1945–95 from a four-member ensemble of GHG+SUL forced runs from the HadCM2 model, but with slight change in color bar from that used in Fig. 7. (top) Ratio of ensemble average 1945–95 temperature change to the standard deviation between the four independent members of the HadCM2 ensemble. This ratio is referred to as the S/N ratio. The gray area shows where the S/N is less than or equal to 1.

Both the model predictions and observations are projected onto this attribution subspace, an action that is akin to projecting different datasets onto a common basis set such as in standard EOF analysis. The uncertainties in the model and data are represented in this space by probability distributions characterized by elliptical confidence regions (in our later examples in a 2D space). In particular, the natural climate variability is represented by an ellipse centered on the origin. If the model signal falls outside the 95% confidence ellipse, say, of the natural variability, one states that a climate change signal due to other than natural processes has been detected at a significance level of 5%, that is, the probability that a signal as large as the observed climate change occurs due to natural climate variability is less than 5%. This does not necessarily imply that the observed climate change can

be attributed to an assumed forcing mechanism. However, if the climate change signal falls within the overlap region of the confidence regions of the model-predicted signal and the data errors, one can state further that the climate change signal is *consistent* with the model's forcing mechanism(s). But this still does not imply that the signal can be uniquely attributed to the hypothesized mechanism, since it is conceivable (and in fact often the case) that other forcing mechanisms could also satisfy the consistency criterion. Unique attribution can be claimed only if it can be demonstrated that there exist no forcing mechanisms, other than the assumed mechanism, that are consistent with the data. Since this is clearly impossible in a rigorous sense (there can always exist mechanisms one has overlooked), detection and attribution claims necessarily have the nature of statistical consistency statements.

c. Bayesian approach to detection

The Bayesian approach (Hasselmann 1998; Leroy 1998) starts from an assumed subjective a priori probability (the "prior") for the existence of an anthropogenic climate change signal. The analysis of the data then provides new information that modifies the prior, yielding an a posteriori probability (the "posterior"). The relation between the prior and the posterior depends on the relative likelihoods of the outcome of the data analysis, given that the hypothesis of an anthropogenic climate change is either true or false. The relation ("Bayes' theorem") follows from straightforward application of the rules of conditional probabilities (see appendix B). Conceptually, the advantage of the Bayes approach is that it considers not only the probability of the validity of the null hypothesis (e.g., climate change due to natural variability), as in conventional theory, but also the probability of the validity of the complementary climate change hypothesis, for the rejection of one hypothesis in favor of the other must clearly depend on the priors of both hypotheses.

In practice, the main advantages of the Bayesian formalism over the conventional approach is that it enables a number of different climate change indices, whose noise level estimates may be associated with very different levels of uncertainty, to be incorporated into a single comprehensive test that exploits all available information. This generally yields, if the indices are statistically independent, much enhanced significance levels compared with a conventional detection and attribution test based on only a few variables, such as near-surface temperatures, for which adequate statistics exist. Another, more formal, advantage is that there is no longer any need to distinguish between detection and attribution, as in the conventional approach; both issues can be combined into a single hypothesis test.

In summary, the Bayesian approach allows a wide variety of information to be used in the detection analysis. The analysis, being subjective, also allows an equally wide set of beliefs to be incorporated into the detection formalism. As the example of appendix B shows, very different sets of initial beliefs might nevertheless be modified by cumulative observed climate change information to converge on a closer agreement as to the causes of observed climate change than might have been anticipated.

5. Detection results

This section combines the results of the previous sections to derive some general conclusions on detection and attribution of model-generated anthropogenic signals in observations of SAT, specifically the temperature trends over approximately the last 50 years for June–August. Before applying the various methodologies, it is informative to first visually compare the model-predicted and observed temperature changes. We then explore the implications for detection and attribution using the optimal fingerprint method with a number of different models, taking into account the various forms of uncertainty discussed above. Subsequently we consider the impact of removing the global mean from the analysis. The list of models used in this section, with their resolution and type of forcing, is shown in Table 2. The analysis is restricted to the data-adequate regions of the earth, but one could also work with selected stations or grid boxes as have Barnett (1986), North and Stevens (1998), and others; the choice is a matter of taste.

a. Visual comparison

Spatial patterns of summertime temperature trend over the period 1946–95[3] predicted by 11 different realizations from three different models forced by GHG+SUL were averaged together and compared in

[3] The end year 1995 was chosen to avoid impacts of the recent large ENSO activity on trend estimation. Results including data through 1998 are similar.

Fig. 10 with the corresponding trend estimates from observations (Jones et al. 1999). The observations show warming over much of the Southern Hemisphere oceans (as far as data exist), and much of the Northern Hemisphere landmass (except eastern Asia), while cooling trends occur over most of the Northern Hemisphere oceans. In contrast, most models simulate rather weaker warming over the Northern Hemisphere landmasses than observed and miss the cooling of the Northern Hemisphere oceans. Also, while the observations show strong hemispheric asymmetry in the warming, it is less apparent in the average model signal. The average of the models appears visually to underestimate the observed land–sea temperature contrast, a key factor in detection (North and Stevens 1998). However, the individual model ensembles do differ on some of these features, for example, the MPI runs where the stronger Southern Hemispheric warming may be partly due to model drift. Note also that the average of all the model runs shows less influence of internal variability (less noise) and hence smaller contrasts than the individual model responses or the single observed trend pattern.

In summary, visual inspection shows many similarities between model simulations and observation but also some substantial differences. In short, simple visual inspection does not allow a clear statement as to whether the model signals are present in the observations.

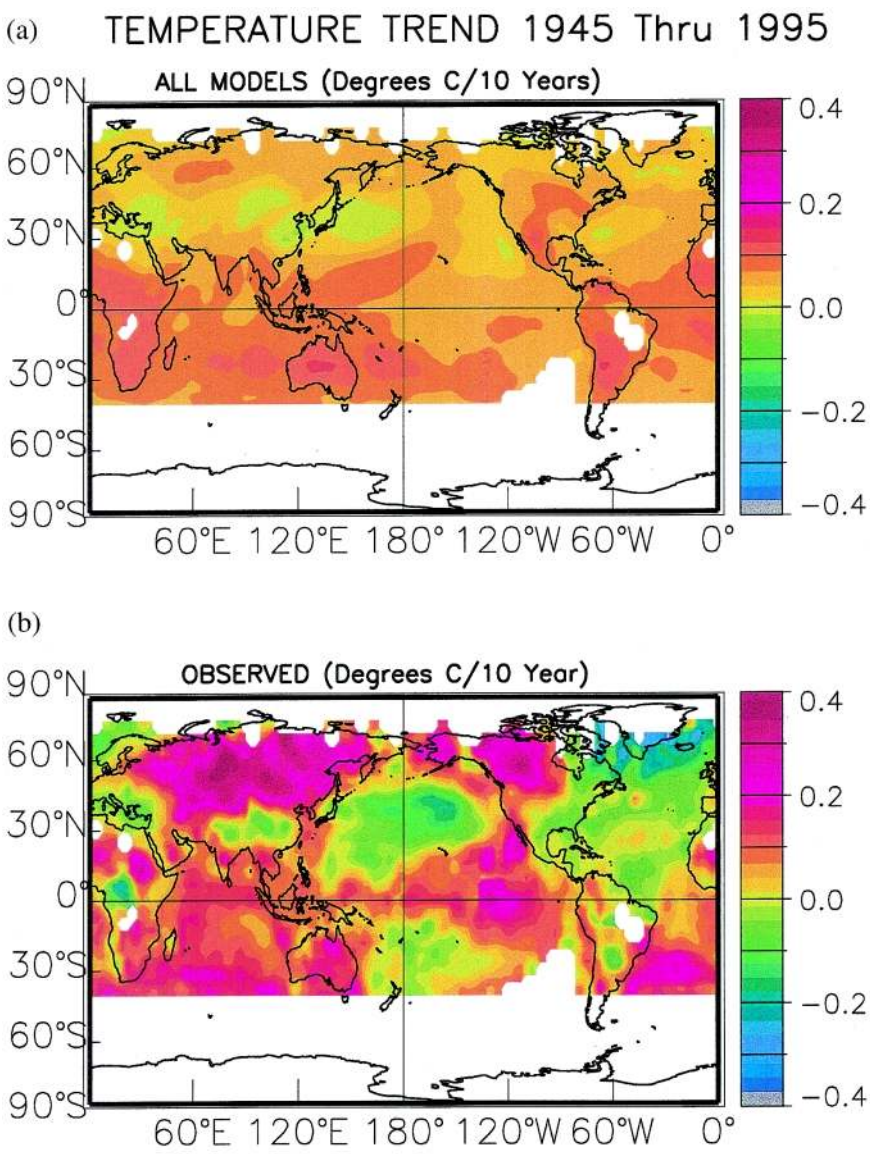


FIG. 10. (top) Temperature trend for JJA between 1945 and 1995 obtained by averaging together the ensembles from HadCM2, GFDL, and MPI. All models were forced by their independent estimates of greenhouse gases and direct sulfate aerosol effects. (bottom) Observed temperature trend for JAA between 1945 and 1995 from Jones et al. (1999).

b. Optimal fingerprint detection

The optimal fingerprints estimated according to appendix A from the MPI's ECHAM3/LSG model and the Hadley Centre's HadCM2 model were used independently as anthropogenic basis sets ("fingerprints") on which to project the observations. Different estimates of the anthropogenic forcing and their uncertainties (section 3) were also projected onto these coordinate systems. Thus, many different model runs, their uncertainties, and the observations can all be compared in a common coordinate system.

The simplest coordinate system is the one-dimensional system defined by a single response pattern to combined GHG and SUL forcing. We designate this as the "GS fingerprint." The one-dimensional detection analysis using a single GS fingerprint is aimed at quantifying an anthropogenic signal in observations and comparing it with the presently expected *magnitudes*

of model signals. Additionally, a GHG and a SUL fingerprint are applied synchronously in order to separate the response to greenhouse gases from that to sulfate aerosols. This two-dimensional fingerprint regression approach was found to be more suitable to distinguish between forcing hypotheses (Hegerl et al. 1997). In this latter coordinate system, the abscissa denotes the amplitude of a pure greenhouse gas pattern and the ordinate the orthogonal sulfate induced pattern (i.e., the sulfate pattern after subtraction of the component parallel to the greenhouse pattern). We shall refer to this 2D space as *attribution space*. The results described below are discussed in more detail by Hegerl et al. (1999).

1) Results: GS fingerprint

The one-dimensional analysis (Fig. 11a) shows that, when projected onto a GS fingerprint, independent of whether the HadCM2 (dashed vertical lines) or ECHAM3/LSG (solid vertical lines) fingerprints are used, the observations (labeled "1" on the abscissa and denoted by a black dot) represent significant climate change at more than the 5% significance level, this level being estimated from a *t* test. That is, these estimates do not include zero and so are not likely to be due to natural variability. The 5% significance level reflects uncertainty due to internal climate variability in observations and due to differences in gridbox value sampling between observations and model [G. C. Hegerl et al. 1999b; for details of the analysis, see Hegerl et al. (1997) and (1999a)]. HadCM2 has been used to estimate the internal climate variability, since it provides the *largest* standard deviation of internal variability in the direction of the GS fingerprint and so yields the most conservative answer. Increasing the level of internal variability by a factor of 2 reduces the significance level to 10%.

Other results demonstrated in Fig. 11a are as follows.

1) Abscissa values 2–5 correspond to ensemble-averaged GHG+SUL runs from HAM3L, HadCM2, GFDL, and HAM4P, respectively (see Table 2 for model identification). Both the HAM3L and HadCM2 GHG+SUL signals are consistent with the observations according to the *t* test for either choice of GS fingerprint. The consistency of the GFDL and the HAM4P GS runs with the observations depends on which model fingerprint is used; for example, it is inconsistent (denoted by an "x" on the abscissa origin) with the observations if the

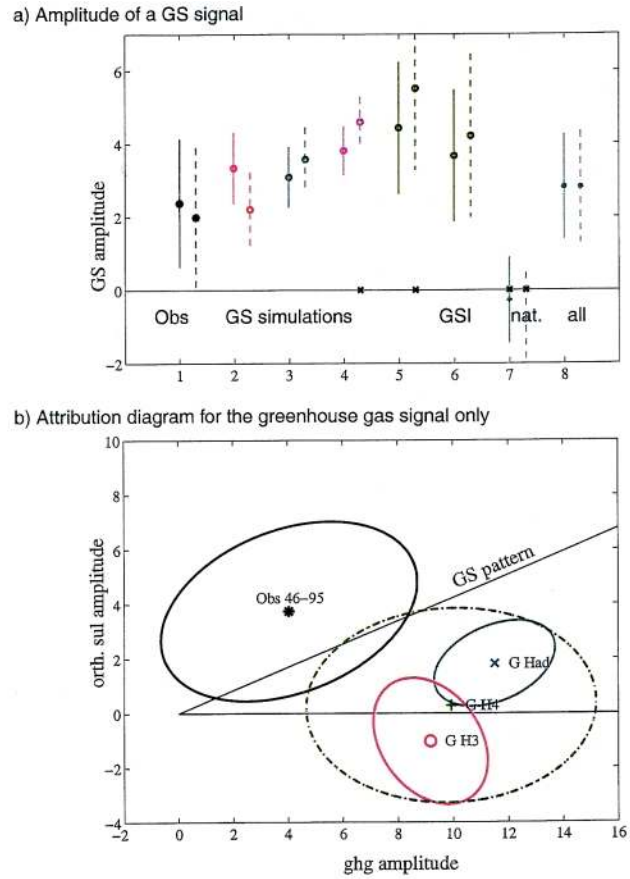


Fig. 11 (a). Model predictions and uncertainties in the amplitude of the one-dimensional GS fingerprint pattern. The type of data/simulation is given by the abscissa value: 1 = observations, 2–5 GHG+SUL forcing, 6 = GHG+SUL+indirect sulfate forcing, 7 = natural forcings (solar/volcanic) from the HadCM2, and 8 = natural and GHG+SUL forcing (all) from HadCM2. The points are the best estimates of the signal strength in GS space. The vertical lines are 90% confidence levels on each signal estimate. Dashed line for GS fingerprint from HadCM2 and solid for HAM3L. The "x" on the abscissa denotes signal estimates that are not consistent with the observations. Colors: black = observations, blue = HadCM2, red = HAM3L, pink = GFDL, and green = HAM4P. See Table 2 for model details and references, and Hegerl et al. (1999a) for details of the calculations and significance tests. (b) Model predictions and associated uncertainty regions for greenhouse gas forcing alone. The points in the attribution space are the location of the various models' estimated trend in surface temperature over the period 1946–95. The model simulations can be compared with the location of the observed temperature trend over the same period (black asterisk). The color code identifies the model, e.g., blue for the average of the HadCM2 integrations, red for Ham3L, and green for Ham4P. See Table 2 for more information on the models. The confidence limits on each simulation are shown by ellipses (dashed for HAM4P). Since there is little or no overlap between model uncertainty ellipses and observed uncertainty ellipses, the observations do not support the hypothesis that greenhouse gases alone can explain the observed change in summer near-surface temperature trend over the last 50 yr. The global mean has been included in the above calculations. For details, see Hegerl et al. (1999a).

HadCM2 fingerprint is used but the reverse is true if the MPI fingerprint is used. So two of the four simulations have an anthropogenic signal whose *magnitude* is consistent with the observations for both sets of fingerprints.

2) Adding indirect sulfate effects puts HAM4P signal (abscissa coordinate 6, GSI defined in Table 2) into agreement with observations.
3) Natural forcing due to volcanoes and solar variability (abscissa = 7, "nat," as simulated by a linear sum of the HadCM2 solar and volcanic forced simulations) projects negatively onto both GS fingerprints and is found to be inconsistent with the observations.
4) Combining both natural and GS forcing ("all") produces a signal in the HadCM2 model that is in good agreement with the observations (Tett et al. 1999; see also North and Stevens 1998; Stevens and North 1996).

2) Results: Attribution space

Separation of the GS signal into its component parts offers a better view of how the model response is composed of a GHG and SUL pattern and hence leads to more powerful attribution statements. So we discuss next the projection of the various model runs, uncertainties, and observations onto attribution space.

(i) Anthropogenic signals

Figures 11b and 12 show estimates for the amplitude of a greenhouse gas fingerprint (GHG) and an orthogonal sulfate aerosol fingerprint (SUL) in observations and model simulations. These results again are based on using 50-yr trend patterns (1946–95) for summer (JJA) surface air temperature, as in the one-dimensional case. The optimal fingerprints are derived from HAM3L signal patterns [see appendix A, Eq. (A2)] and, for Fig. 12b, from HadCM2 signal patterns. Estimates of internal variability in observations are again based on HadCM2 in control simulations in order to be conservative.

Figure 11b shows a comparison between the amplitudes of signals from greenhouse-gas-only simulations from the different models and observations. Note that in this case, any amplitude of an SUL pattern (ordinate values other than zero) in the models would occur only due to internal variability or to model–model differences since the forcing was by GHG alone. The position of the observed 50-yr trend pattern (*) in this space is uncertain due to internal cli-

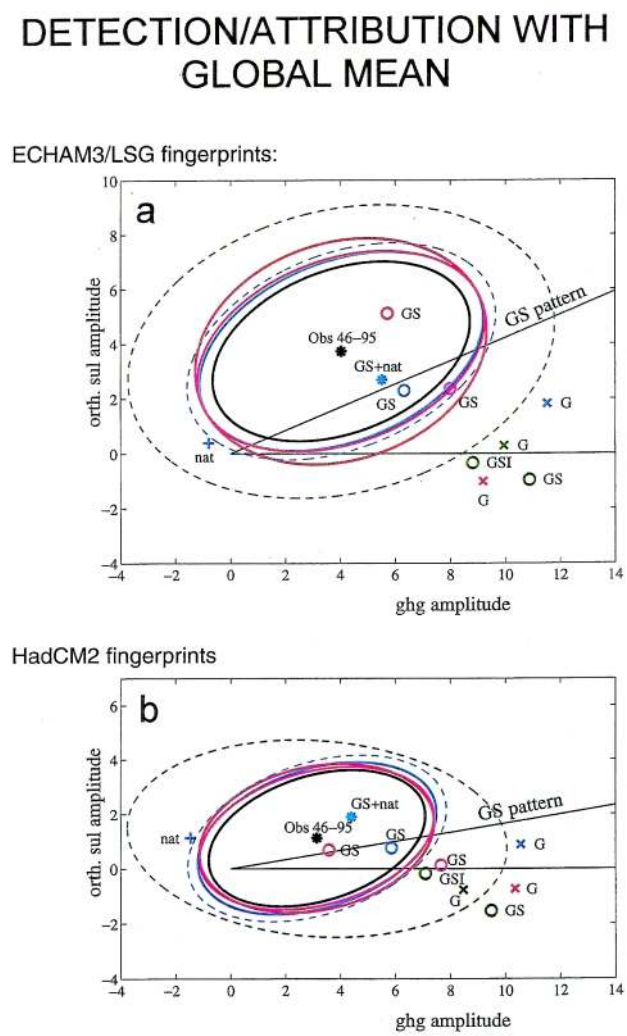


Fig. 12. The same as Fig. 11b. Detection results are presented in attribution space defined by the HAM3L (upper) and HadCM2 (lower) models. A variety of runs using different combinations of anthropogenic and/or estimated natural variability are shown. Red denotes simulations with HAM3L, blue with HadCM2, green with HAM4P, and pink with GFDL. The G denotes greenhouse-gas-only forcing, GS greenhouse gas + direct sulfate aerosol effect, while "nat" refers to a linear combination of solar and volcanic forcing from the HC run. The radiative forcing of the HAM4P simulations (GSI) includes an updated estimate of the direct and indirect sulfate aerosol effect, the major greenhouse gases, and tropospheric ozone (Roeckner et al. 1999).

Model simulations that lie *outside* their respective uncertainty ellipse (which is based on internal variability noise in observations, the ensemble average in the simulation, and the observational sampling error) are inconsistent with the observations at the 5% level. The confidence ellipses have now all been centered on the current observed state. The much larger confidence ellipse (dashed) is associated with the HAM4P run for which there is only one realization. The global mean is included in the climate change signal.

mate variability and sampling error. The position of the model trend estimates (o, +, x) is uncertain due to internal model variability. Both uncertainties are shown as ellipses around their respective best estimates. In Fig. 12, which now includes simulations with a variety of different forcings in addition to GHG alone, both observational and model uncertainty are combined into a joint ellipse for the differences between model and observed climate state, which is centered around the observed climate state. Where the models' points lie outside their respective ellipse, the difference from the observations is too large and the model is inconsistent with the observations at the given significance level.

A comparison of the various model predictions and their projection onto the ECHAM3/LSG and HadCM2 fingerprints leads to the following conclusions.

1) The black ellipse defines the uncertainty region of the observed signal, in the GHG and SUL pattern space, associated with the natural variability and observational sampling error. In the MPI world (Fig. 12a), the ellipse does not contain the origin point (0,0), implying that internal climate variability alone cannot explain the recent observed trend. When the HadCM2 fingerprints are used, however, the joint uncertainty ellipse encompasses the origin, so that an explanation of the observed trend by natural variability cannot be excluded (Fig. 12b). However, this explanation was excluded in the 1D analysis of section 5b(1). Also, application using observed trends through 1998 and a full space–time detection method using the HadCM2 model (Tett et al. 1999; P. A. Stott et al. 1999, manuscript submitted to *Climate Dyn.*) produced a conclusion similar to that obtained in the ECHAM3/LSG attribution space. At this stage, the detection and attribution conclusions clearly are sensitive to and depend on the model fingerprints and detection techniques used in the analysis.
2) The GHG-only simulations from the different models (denoted by G in Figs. 11b and 12) are all substantially inconsistent with the observations; that is, the experiments occupy positions in the attribution space that fall well outside the 95% confidence ellipse of the observations. Thus the simulated temperature trends produced with greenhouse forcing only are not consistent with the observations.
3) The observations (black asterisk in Figs. 11b, 12, and 13) in the attribution space of the ECHAM3/LSG have a significant ordinate component, indicating a substantial influence of the sulfate forcing. This component is somewhat larger in the observations than predicted by all model simulations except MPI GHG+SUL (red circle labeled GS). In the HadCM2 fingerprint, the observations are not significantly separated from the origin, suggesting an insignificant sulfate response, for example, sulfate has no significant impact on SAT in that model. This suggests the detectability of the sulfate response in observed trend patterns is model dependent. Further, Stott et al. (1999, manuscript submitted to *Climate Dyn.*) found explicit inclusion of the time-dependent response to sulfate gives a clear identification. So the identification of this signal also depends on the detection method one uses.
4) The simulations forced by GHG+SUL (GS) are consistent with the observations for two of the four models using either fingerprint. However, as we saw in the 1D case, the GFDL result (pink circle) depends on which fingerprint one uses and is marginally consistent or inconsistent in either case.
5) The HAM4P simulations, which include the most sophisticated, yet novel, chemistry package, lie outside the 95% confidence limits for all experiments (green, G, GS, GSI) in the ECHAM3/LSG fingerprint and the GS experiment on the HadCM2 fingerprint. Notice the weak, almost negligible sulfate response (no significant ordinate values) based on either fingerprint. This feature arises in HAM4P since the combined direct and indirect sulfate signal is moderately uniform over the globe. So it looks like the GHG signal but of opposite sign. Hence it projects onto the GHG axis and the two signals become hard to separate (Feichter et al. 1997). Unfortunately, we have only one realization of this simulation and a short control run, hence the huge uncertainty ellipse, and no comparable simulations from other models to substantiate it. Until these limitations are removed, the validity of this simulation cannot be judged. The danger of basing detection statements, either positive or negative, on a single scenario run is well illustrated in KDDS and BHKT.
6) The HAM4P signal was found to be statistically consistent with the observations if we redid the detection test using 30-yr trends of anually averaged SAT and a fingerprint derived from HAM4P itself (R. Schnur, personal communication). This result illustrates the potentially strong dependence that detection statements have on the fingerprint used,

the time period, and the season considered in the test. These dependencies have not been well explored.

(ii) Natural and anthropogenic forcing

Various types of "natural" forcing have been included in the HadCM2 simulations, either together with or separately from the anthropogenic forcing. The principal results from these additional experiments are as follows.

1) A superposition of simulations forced with solar insolation variations and volcanic forcing ("nat," blue +) is found to be (marginally) inconsistent with the observations. The same conclusion, but at a higher significance level, was found by Tett et al. (1999) using a different detection scheme. Similarly, Hegerl et al. (1997) find that climate variations induced by changes in solar insolation alone (Cubasch et al. 1997) are unlikely to explain the observed temperature increase (not shown). We conclude that natural forcings cannot convincingly explain the most recent temperature trends, unless reasons can be found to increase the magnitude of the solar forcing above current estimates.
2) The combined simulation with the HadCM2 model including natural forcing by solar insolation variations and volcanic activity together with the GHG+SUL anthropogenic forcing (blue asterisk: GS+nat) is in good agreement with the observations, yet very similar to GS forcing alone. This result was obtained also by Tett et al. (1999) using a different methodology and was found to be robust relative to uncertainty in estimates of natural variability. The simulation does not include poorly known, but probably important, radiative forcings such as indirect sulfate effects (Hansen et al. 1998). North and Stevens (1998) and Stevens and North (1996) also find significant GHG, SUL, and volcanic signals in observations applying a space–frequency detection method.

(iii) Summary

The results of the last two sections indicate that greenhouse gas forcing alone is inconsistent with the observations in all the coupled global climate model simulations studied. The anthropogenically forced runs with combined greenhouse gas and sulfate forcing (GHG+SUL) from two of the four models are consistent with the observations; the other two models are either consistent or inconsistent with the observations depending on the fingerprint used (Hegerl et al. 1999).

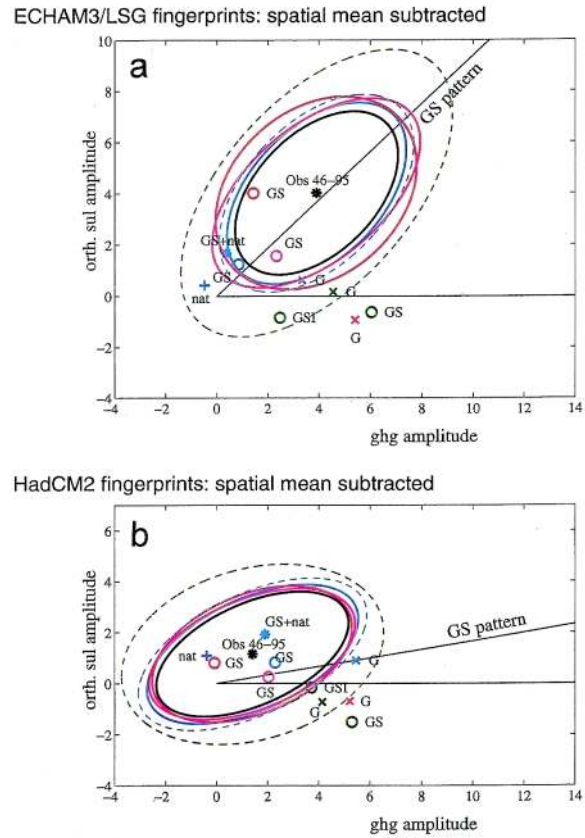


FIG. 13. Same as Fig. 12 except that the global mean has been removed from the signals.

A similar conclusion about consistency of the HadCM2 model has been reached by Tett et al. (1999), using a different detection strategy. However, the MPI model HAM4P with the most complete, yet novel, chemistry is either consistent or inconsistent with observed warming (Roeckner et al. 1999), depending on which fingerprint in 1D or 2D space one uses, the time period studied (50-yr vs 30-yr trends), and seasonality of the signal (summer vs annual average). However, we note this result is based on a single model run (cf. section 3b) and is presently unconfirmed by other models forced in a similar fashion.

Simulations forced with current estimates of natural forcing alone (variations of solar insolation with or without variations of volcanic activity) are found to be inconsistent with observed temperature changes. Simulations forced with both natural and partially complete anthropogenic forcing show rather good agreement. This result needs to be reassessed when simulations with other models are available, and

when a combination of realistic natural and anthropogenic forcings are included in improved model simulations.

c. Detection without the global mean

The analyses described above included temporal changes both in the global-mean temperature and in the spatial variations about the global mean over the last 50 years. The change in the global-mean temperature, while clearly a sensitive indicator of global warming, is obviously not a very useful quantity for discriminating between competing mechanisms of climate change. For the purposes of *attribution,* it has been suggested that it would be more effective to remove the global mean, thus emphasizing the differences between the responses to different forcing mechanisms (e.g., Barnett 1986).

The impact of removing the temporal changes in the global mean is shown in Fig. 13, which may be compared with the corresponding Fig. 12 with the global mean included (see also Hegerl et al. 1999b). Most of the above conclusions are seen to remain valid after exclusion of the global mean. Indeed, some model runs that were inconsistent with the observations now come into closer agreement with the observations. Note that in the HadCM2 world, the observational uncertainty still includes the origin in either case; compare Fig. 12b and Fig. 13b.

d. Summary

The detection of time-dependent patterns of climate change associated with anthropogenic forcing gives much the same results with and without inclusion of the global mean. This suggests that the detection scheme used here relies for its skill not only on changes in the global-mean temperature, but to a comparable or greater part on changes in the spatial patterns of temperature change. This is an encouraging result for the application of fingerprint techniques to the problem of attribution.

6. Summary statement on detection and attribution

Based on our analyses described in section 5, we summarize the present situation regarding the detection and attribution of an anthropogenic climate change signal as follows.

The recent changes in global climate inferred from near-surface atmospheric temperatures cannot be readily explained by natural climate variability. However, the significance level at which the null hypothesis is rejected depends on the detection methodology used (cf. Fig. 12).

Greenhouse warming alone is insufficient to explain the observed pattern of climate change. Most models reproduce the observed temperature trend patterns better if the direct effects of sulfate aerosols are included with GHG. Nevertheless, it is worth noting that the predicted temperature change pattern of the model with the most advanced chemistry, including other poorly known but potentially important forcings, was either consistent or inconsistent with the observations in 2D attribution space depending on which fingerprint, time period, and season were used for detection. As noted above, this single run is unique and we have no corroborating evidence, currently, with which to either accept or reject the model's main result. In any event, recent results have shown it can be misleading to base detection statements on a single scenario run. Such statements also depend on what method, fingerprints, etc. are used.

The most probable cause of the observed warming is a combination of internally and externally forced natural variability and anthropogenic sources (see also Tett et al. 1999). But given the large model uncertainties and limited data, a reliable weighting of the different factors contributing to the observed climate change cannot currently be given. In short, we cannot attribute, at this time, with a high level of statistical significance, the observed changes in global and large-scale regional climate to anthropogenic forcing alone.

This result should not come as a great surprise. Although the results shown in Fig. 12 suggested attribution to anthropogenic forcing could be made, the simulations showed considerable scatter and neglected important factors such as the indirect sulfate effects, ozone, etc.; and the single model run that included these poorly known effects was not conclusively consistent with the observations, a result that cannot be corroborated until more realizations from that and other similarly forced models are available for analysis.

The current situation is that a fully realistic ensemble of scenario runs has not yet been conducted. This is compounded by the difficulty that, by most estimates, the anthropogenic signal is currently comparable in magnitude to the upper limits of the natural climate noise. Such a low signal to noise ratio makes clear attribution statements difficult at this time. While

some scenario runs by some models reproduce the observations to within the level of estimated uncertainty in the signal and observations, others do not. In short, the current state of affairs is not satisfactory.

7. Recommendations for future work

Sustained efforts in several research areas are required to improve the situation. On the basis of our analysis, we identify the following as most urgent.

1) An ensemble of scenario simulations for different anthropogenic and natural forcings, using realistic climate models, including the key chemical species, is central for further work. These runs need to be intercompared [e.g., within the framework of the planned Coupled Model Intercomparison Project; Meehl et al. (1997); Gates et al. (1999)] and made widely available to a large number of users.
2) Multivariate detection studies are required that include all available climate change indices, such as modifications in the annual and diurnal cycle, vertical temperature structure, precipitation, sea ice extent, glacier retreat, and sea level. These studies also need to include the statistics of natural variability, such as ENSO, the North Atlantic oscillation, and other large-scale oscillation modes, as well as changes in the probability distributions of shorter-term weather phenomena. Models will be needed both to predict the relevant climate change signals and to estimate the natural variability levels. Since the model estimates will be unavoidably contaminated by model errors, many of which will be difficult to quantify, a Bayesian approach may be needed to combine all information into a single comprehensive detection and attribution analysis.
3) Existing detection and attribution theory should be applied to make use of not only the predicted spatial structure but also the full time dependence of the multivariate signal fields as suggested by Hasselmann (1993). This would largely remove any dependence on the results on specific time intervals (such as the 50-yr trends used in this work) or seasonal character of the signal (summer vs annual average temperatures). A first step in this direction has been taken by Tett et al. (1999) and North and Stevens (1998). Further, new detection paradigms need to be explored and developed.
4) The basic datasets required for detection need to be upgraded. A reconstruction of the history of the three-dimensional atmospheric temperature field derived from radiosondes, after correction for various biases, would be a good place to start. It is noteworthy that several nations have begun such programs. Continued upgrading of the near-surface temperature field is also required. In addition, further datasets for detection need to be developed.
5) The problems with the paleodatasets require immediate attention. In the final analysis, they are our only means to determine levels of natural variability over the last 1000 or so years from real data.
6) In the process of extending the datasets for detection and attribution, it would be useful to consider data that are perhaps not optimal for early detection and attribution, but are nonetheless of considerable public interest regarding their possible modification through anthropogenic greenhouse warming. Thus statements on whether and when changes can be detected in the frequency of occurrence of hurricanes, midlatitude storms, floods, droughts, El Niño, and other extreme weather or short-term climate events would be of considerable practical value. This would be true even if it should be found that these climate change indices are emerging more slowly from the natural variability noise than other climate change indices, such as the global near-surface temperature pattern.
7) While obvious, there is a continuing need to upgrade the climate models used to make the anthropogenic predictions. All of the models discussed here can improve in the areas of simulating natural variability, ocean processes (especially mixing), cloud representation, etc.

Acknowledgments. The coordinated approach to this work was made possible by NOAA's Office of Global Programs and the Department of Energy's Office of Energy Research in conjunction with the Climate Change Data and Detection element (a part of the U.S. Global Change Research Program). The Scripps Institution of Oceanography provided partial support for Tim Barnett. Gabriele Hegerl was supported by the Alexander von Humboldt Foundation, NSF, and University of Washington through JISAO. S. F. B. Tett and computer time for the HadCM2 simulations were supported by the U.K. DETR under contract PECD 7/12/37.

The model data were provided as follows: The HadCM2 runs have been supplied by the Climate Impacts LINK project (U.K. Department of Environment, Transport and the Regions Contract EPG 1/1/16) on behalf of the Hadley Centre and U.K. Meteorological Office. The MPI runs were provided by the Germany Climate Computing Center (DKRZ), courtesy of L. Bengtsson, E. Roeckner, and U. Cubasch. R. Schnur (MPI) was a great help

with data issues. The Geophysical Fluid Dynamics Laboratory was especially generous with their data since runs were provided while they were in progress. The SAT dataset used here has been developed jointly by Phil Jones and David Parker.

Special thanks are due to Phil Offen for managing to pull together the many different components of this report, particularly in view of the many different computers and word processing systems used. Tony Tubbs, David Pierce, Marguerite Schultz, and Niklas Schneider (all SIO) provided critical help in different areas of manuscript preparation. We also thank Keith Briffa and Tim Osborn (UEA) for assisting with collection and analysis of the paleo–proxy data series.

Appendix A: The conventional formalism

a. Optimal fingerprints

A standard assumption of all detection and attribution methods is that the observed climate change vector $\boldsymbol{\varphi}(t)$ can be represented to first order as a linear superposition,

$$\boldsymbol{\varphi}(t) = a(t)\mathbf{g}(t) + \mathbf{n}(t) \tag{A1}$$

of the signal $a\mathbf{g}$ and noise $\mathbf{n}$, where the signal is decomposed into an amplitude a and a suitably normalized signal pattern $\mathbf{g}$. This notation applies to a single pattern case and can easily be generalized to multiple patterns. The climate change signal pattern $\mathbf{g}(t)$, observed climate change vector $\boldsymbol{\varphi}(t)$, and climate variability noise $\mathbf{n}(t)$ are defined here in the space of some set of climate variables (temperature, pressure, precipitation, etc.) for which individual or grid-averaged measurements exist in some spatial distribution.

In both the conventional and Bayesian approach, one wishes to extract from the data an estimate of the climate change signal and a measure of the signal-to-noise ratio. This is normally achieved by applying a suitable linear filter $\mathbf{f}(t)$ to the data (Hasselmann 1979, 1993, 1997, 1998; Bell 1982, 1986; North et al. 1995; North and Kim 1995; Hegerl et al. 1996, 1997; Santer et al. 1996a; Allen and Tett 1999). Where applicable, the filter is designed to yield a detection variable d that maximizes the signal-to-noise ratio. Thus, the goal is to design a filter $\mathbf{f}(t)$ that projects the observed data (defined in the interval $(t_0 \le t \le t_1)$ onto a detection variable, d, that is,

$$d = \int_{t_0}^{t_1} \mathbf{f}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t)dt, \tag{A2}$$

(using matrix notation, the index T denoting the transpose), such that the square signal-to-noise ratio

$$S = d_g^2 \,/\, \langle d_n^2 \rangle \tag{A3}$$

is maximized. Here d_g denotes the detection variable for the case of a pure signal in the absence of noise, $\boldsymbol{\varphi}(t) = \mathbf{g}(t)$, and d_n is the detection variable computed for the complementary case of pure noise in the absence of a signal, $\boldsymbol{\varphi}(t) = \mathbf{n}(t)$. Cornered parentheses denote expectation values. Maximizing S is also equivalent to maximizing the square ratio $\langle \mathbf{d}^2 \rangle / \langle \mathbf{d}_n^2 \rangle$ of signal plus noise to noise.

Equation (A2) can be written in the more compact notation

$$d = \mathbf{f}^{\mathrm{T}}\boldsymbol{\varphi} \tag{A4}$$

by combining the discretized time variable with the component index denoting the location (and, if relevant, type) of the climate state vector component in a single composite vector-component index. Then, $\boldsymbol{\varphi}$ denotes the climate state trajectory.

Instead of integrating over the full time interval, yielding a time-independent net detection variable d, it is often convenient to define a time-dependent detection variable $d(t)$, using a running filter (representing, e.g., a trend estimate over a running time interval). This is the approach used in the examples discussed in section 5.

It is important to note that the optimal filter differs from the predicted climate change signal $\mathbf{g}(x,t)$. Although the signal pattern is sometimes used as a detection filter, this is not the optimal choice, as it ignores the properties of the natural variability noise, characterized by the covariance matrix $\mathbf{C} = \langle \mathbf{n}\,\mathbf{n}^{\mathrm{T}} \rangle$, which one wishes to suppress relative to the signal. Mathematically, the signal and filter patterns represent co- and contravariant duals with respect to a metric given by the inverse of the covariance matrix $\mathbf{C}$ (Hasselmann 1979, 1993, 1998).

b. Estimating the signal amplitude

When inferring the climate change signal from the observations, the number of degrees of freedom of the estimated signal must be strongly reduced by requiring the signal pattern to lie in some low-dimensional pattern space. This is essential for a successful detection and attribution strategy. Otherwise the signal is

swamped by the many degrees of freedom of the noise (cf. Barnett and Preisendorfer 1987; Hasselmann 1993, 1997). In the simplest case that only a single climate change mechanism is being tested, the number of degrees of freedom can be reduced to one by specifying the direction of the extracted signal to agree with the predicted signal pattern. Thus the only free parameter inferred from the observations is the signal amplitude. In most DA applications, however, a number of candidate climate change forcing mechanisms are considered simultaneously and a multipattern analysis is called for. For brevity, we summarize in the following only the single pattern case; the multipattern generalization is straightforward.

We define as the best estimate $\hat{a}$ of the signal amplitude the value for which the mean square of the residual $\mathbf{r} = \boldsymbol{\varphi} - \hat{a}\mathbf{g}$ between the observed climate change and the estimated signal is minimized with respect to the metric $\mathbf{C}^{-1}$:

$$\langle \mathbf{r}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{r} \rangle = \min. \tag{A5}$$

Minimizing the residual according to (A5) is equivalent to maximizing the probability (for Gaussian variability) that the residual can be attributed to the natural climate variability and therefore contains no external forcing signal.

If the amplitude $\hat{a}$ is estimated from the observations by some linear process, that is, by the application of a filter $\hat{\mathbf{f}}$,

$$\hat{a} = \hat{\mathbf{f}}^{\mathrm{T}} \boldsymbol{\varphi}, \tag{A6}$$

the optimal filter that minimizes the mean square error (Hasselmann 1998) is found to be

$$\hat{\mathbf{f}} = \mathbf{f}, \tag{A7}$$

so that

$$\hat{a} = d. \tag{A8}$$

Thus the detection variable d represents also the optimal estimate of the signal amplitude.

c. Confidence levels

As mentioned in the main text, in the conventional approach the detection and attribution problems are separated. First, one determines whether the detection variable d exceeds some critical value d_c associated with some given probability, $c = 5\%$, say, that d can still be attributed to the natural variability noise. For Gaussian noise, for example, $d_{5\%} \approx 2\sigma_d$, where

$$\sigma_d = \sqrt{\langle d_n^2 \rangle}.$$

If $d > d_c$, the null hypothesis is rejected at the risk level c, and a climate change signal is said to be detected at the significance level $1 - c$ (95%).

If a climate change signal in a given signal direction has been detected at the prescribed significance level, one addresses then the attribution problem. For the single-hypothesis case discussed here, this reduces to determining whether the error $\varepsilon = a - \hat{a}$ between the predicted and inferred climate change amplitude lies within some given confidence interval corresponding to the errors involved in the estimation of the signal amplitude. These consist of two contributions: statistical errors associated with the natural variability noise, and model errors. The statistical errors can be expressed in terms of the covariance matrix $\mathbf{C}$. The model errors, however, cannot be characterized within the framework of conventional statistics, as there exists no adequate model ensemble over which statistical averages can be formed. Conventional attribution tests have therefore normally been based only on the statistical estimation errors of the signal amplitude (cf. Hegerl et al. 1999a).

Appendix B: The Bayesian formalism

In the framework of confirmation theory relevant for our problem, Bayesian statistics is concerned with subjective probabilities of the validity of hypotheses. Prior to the performance of some test E, a person assigns some probability $p(H)$ (the "prior") to the validity ($H = h$ = "true") or invalidity ($H = \bar{h}$ = "false") of a hypothesis H. The outcome $E = e$ (positive) or $E = \bar{e}$ (negative) of the hypothesis test E changes the prior probability into the posterior probability $p(H/E)$. The relation between the prior and posterior probabilities is given by Bayes's basic theorem

$$\frac{p(H/E)}{p(H)} = \frac{p(E/H)}{p(E)}, \tag{B1}$$

which follows directly from the relevant conditional probability definitions. For the case that the hypothesis H is correct, $H = h$, and the test outcome is positive, $E = e$, Eq. (B1) may be written

$$c = c_0 l / \{c_0 l + (1 - c_0)\hat{l}\} \quad \text{(B2)}$$
$$= (1 + \beta \hat{l}/l)^{-1},$$

where

$$\beta = \frac{1 - c_0}{c_0}. \quad \text{(B3)}$$

We have introduced here the shorter notation $c = p(h/e)$ (for "credibility") and $c_0 = p(h)$ for the posterior and prior probabilities, respectively, that the hypothesis is true; $l = p(e/h)$ for the likelihood of a positive outcome of the test for the case that the hypothesis is true; and in place of $p(e)$ the complementary likelihood $\hat{l} = p(e/\bar{h})$ (corresponding to the conventional null-hypothesis statistic of a positive test outcome for the case that the hypothesis is false), where

$$p(e) = p(e/h)p(h) + p(e/\bar{h})p(\bar{h}) \quad \text{(B4)}$$
$$= lc_0 + \hat{l}(1 - c_0).$$

The dependence of the posterior probability c on the prior probability c_0 and likelihood ratio $\hat{l}/l$ is shown in Fig. B1a (see also discussion below).

a. The detection test

The difference between the conventional and Bayesian approach can be illustrated by the standard detection test. The hypothesis in this case is that there exists a climate change signal in the data that cannot be attributed to natural climate variability. The detection test E involves determining whether the amplitude of the climate change signal inferred from observations (for a given signal pattern) exceeds the threshold for which the probability $\hat{l}$ that the amplitude can be attributed to the natural variability noise falls below some prescribed value (say, 5%). If this is the case, $E = e$ = "true," the climate change signal is said to be detected at a significance level

$$a = 1 - \hat{l} \text{ (e.g., 95\%)}. \quad \text{(B5)}$$

This may be compared with the Bayesian expression (Fig. B1a), which can be approximated for small $\hat{l}$ by

$$c \approx 1 - \beta \hat{l}/l. \quad \text{(B6)}$$

The relations (B5) and (B6) differ by the factor β/l,

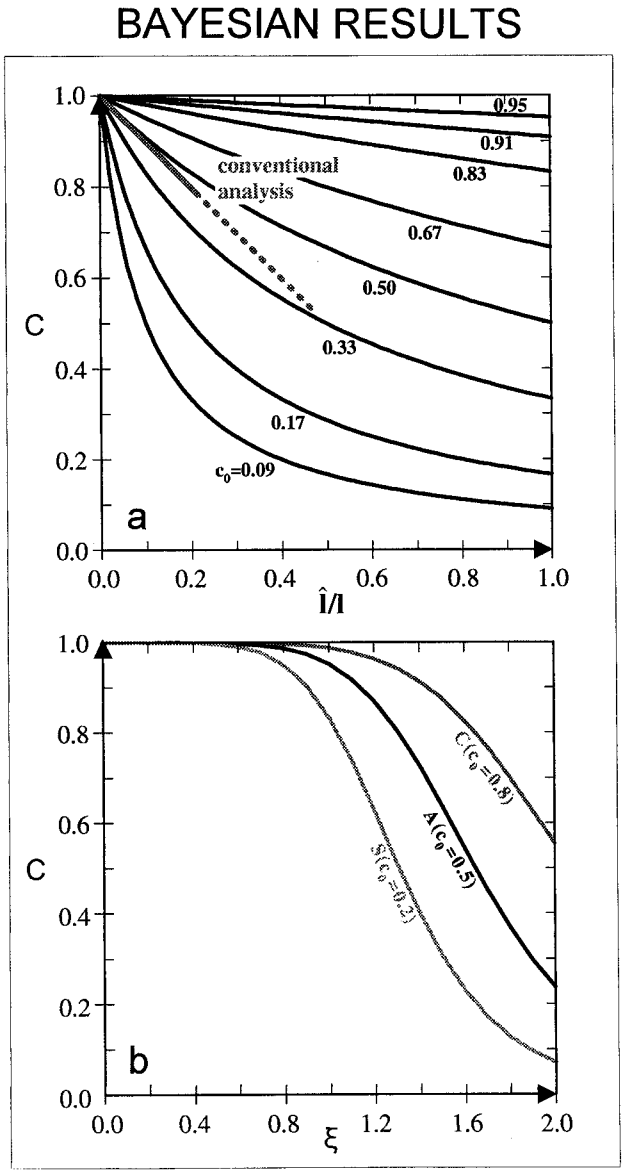


FIG. B1. (a) Posterior probability (credibility) c that a (greenhouse warming) hypothesis H is true given a positive outcome of a detection test. The dependence of c is shown as a function of the assumed prior for c_0 and the likelihood ratio $\hat{l}/l$. Also shown is the line $c = 1 - \hat{l}/l$ corresponding to the conventional non-Bayesian significance level $c = 1 - \hat{l}$ for $l \approx 1$ (from Hasselmann 1998). See text and appendix B for details. (b) Cumulative posterior probability c of greenhouse warming hypothesis H for the six climate change indices defined in Hasselmann (1998) as a function of an index of increasing likelihood that index changes were due to natural variability. The three curves correspond to the three prior probabilities $c_0 = 0.2$ (S, for skeptical), 0.5 (A, for ambivalent), and 0.8 (C, for convinced).

which can adopt any value between zero and infinity, depending on the values of the probabilities c_0 [which determine β through (B3)] and l.

To estimate l one needs to consider the amplitude

of the predicted climate change signal. Thus, in contrast to the conventional approach, in which the signal amplitude is irrelevant for detection and is invoked only when the analysis is extended to the attribution problem (cf. Hasselmann 1997), for the Bayesian analysis it is not possible to separate formally between detection and attribution, the estimate of l involving already an element of attribution.

b. A combined detection/attribution test

A basic shortcoming of the conventional detection test is that it is not symmetrical with respect to the case $H = h$ that the hypothesis is valid and the alternative case, $H = \bar{h}$, that the null hypothesis (or, generally, the complementary hypothesis representing the set of all plausible alternatives to H) is valid. The same limitation applies to the conventional attribution test. These shortcomings are overcome in a Bayesian approach that combines detection and attribution. To distinguish between several competing climate change mechanisms, a general multipattern analysis is required. The inferred and predicted climate change signals are represented as vectors in a low-dimensional space spanned by the predicted signal patterns of the competing forcing mechanisms. We consider then the probability of the signal lying within a small (in the limit, infinitesimal) region of this space at the location of the retrieved signal for each of the competing climate change hypotheses. The probability of a positive outcome $E = e$ of this test for any given hypothesis is infinitesimally small, which would rule out such a test in conventional statistics. However, the test is meaningful in the Bayesian framework, as the impact of the test on the posterior probability c depends only on likelihood ratios, not on absolute likelihoods [Eq. (B2)]. In this manner, we can test and intercompare the relative posterior probabilities for each of the competing climate change hypotheses, including the null hypothesis. Apart from the different test definition, the theory and general relations summarized in Fig. B1 apply exactly as in the detection case discussed in the previous section.

c. Application to multiple climate change indices

As an example, consider the impact of a set of different climate change indices x_i on the hypothesis that there exists already today a greenhouse warming signal. The advantage of combining several different climate change indices is that, even though the information associated with any single index may be uncertain and inconclusive, the cumulative information derived from a cluster of independent observations greatly exceeds the information obtained from a single climate variable.

To test the greenhouse warming hypothesis, we compare the likelihood l of measuring the observed indices for the case that the greenhouse warming hypothesis is true with the likelihood $\hat{l}$ for the complementary hypothesis that the observed climate change is due to natural variability or some other mechanism. Both likelihoods refer to the (infinitesimal) probabilities for the case e that all indices x_i lie in an infinitesimal region of the index phase space located at the observed values. The impact of the observations on the credibility of the greenhouse warming hypothesis is then determined by the probability density ratio [Eq. (B2)]. Assume, for simplicity, that the indices are statistically independent. This will normally not be the case, but the indices can be readily transformed to orthogonal variables. For statistically independent variables, the relevant net probabilities are given by the products

$$l(e) = \mathbf{\Pi}_i \hat{l}_i(e_i), \tag{B7}$$

$$\hat{\boldsymbol{l}}(e) = \mathbf{\Pi}_i \hat{l}_i(e_i), \tag{B8}$$

of the relevant one-dimensional probabilities $l_i(e_i)$, $\hat{l}(e_i)$, for the cases e_i that the individual variables x_i lie within infinitesimal intervals containing the observed values. From Eq. (B2) we obtain then

$$c = \left\{1 + \beta \mathbf{\Pi}_i (\hat{l}_i / l_i)\right\}^{-1}. \tag{B9}$$

Through the product forms (B7)–(B9), the posterior probability of the greenhouse warming hypothesis can become quite high, even when the impact of individual indices remains relatively low.

An example of how the Bayesian approach works is illustrated by a comparison of the single- and multiple-index case in Fig. B1 (adapted from Hasselmann 1998). The single-index case (Fig. B1a) shows the dependence of the posterior c on the likelihood ratio $\hat{l}/l$, where $\hat{l}$ denotes the likelihood that the observed signal lies outside the 95% natural variability confidence ellipse for the case that the null hypothesis is valid and l is the likelihood of satisfying the same test for the complementary hypothesis that there exists a climate change signal. The test clearly depends on the prior (c_0), indicated as a parameter on the curves, that is, on how high one judges the probability in advance that

an anthropogenic signal exists. The corresponding result for the conventional test is also indicated. The considerable differences between the conventional detection test, which is independent of the prior, and the Bayesian result illustrates the marked impact that different prior (subjective) assessments of the likelihood of the existence of an anthropogenic signal can have on DA. Much of the current debate on climate change detection can be attributed to these different priors.

The cumulative impact of considering a number of different climate change indices is illustrated in Fig. B1b, based on an example of Hasselmann (1998). A set of six climate indices was considered: trends in near-surface, midtropospheric, and stratospheric temperature; trends in Arctic and Antarctic sea ice; and changes in the vertical tropospheric profile. The complementary (null) hypothesis to the greenhouse warming hypothesis was again that the climate change indices could be attributed to natural variability. The Bayesian approach was applied for three different priors: $c_0 = 0.2$ (skeptic, S); $c_0 = 0.5$ (ambivalent, A), and $c_0 = 0.8$ (convinced, C). The abscissa axis represents a common factor $\xi = r_i/r_i^{\text{ref}}$ that was applied to the different likelihood ratios $r_i^{\text{ref}} = \hat{l}_i/l_i$ assumed by Hasselmann (1998) for the six individual climate change indices i = 1,2, . . . 6, which yielded a net (subjective) likelihood ratio $r^{\text{ref}} = \mathbf{\Pi} r_i^{\text{ref}} = 0.05$. Thus the variable ξ corresponds to the multivariate-index equivalent of the abscissa likelihood ratio shown in Fig. B1a.

A comparison of Figs B1a and B1b demonstrates the great enhancement in detection and attribution credibility achieved through the use of multiple climate change indices if the indices are statistically independent. (In the illustrative example shown, this was presumably not the case, but the variables can be readily orthoganized through a suitable linear transformation.) The credibility curves are much higher in Fig. B1b, and the credibility curves S and C for the skeptical and convinced observers, respectively, are no longer very different, even though the subjective prior probabilities lie at the two ends of the scale, differing by a factor of 4.

References

Allen, M. R., and S. F. B. Tett, 1999: Checking for model consistency in optimal fingerprinting. *Climate Dyn.,* **15,** 419–434.

Barnett, T. P., 1984: Long term trends in surface temperature over the oceans. *Mon. Wea. Rev.,* **112,** 303–312.

——, 1986: Detection of changes in the global troposphere temperature field induced by greenhouse gases. *J. Geophys. Res.,* **91,** 6659–6667.

——, 1995: Monte Carlo climate forecasts. *J. Climate,* **8,** 1005–1022.

——, 1999: Comparison of near-surface air temperature variability in 11 coupled global climate models. *J. Climate,* **12,** 511–518.

——, and R. Preisendorfer, 1987: Origins and levels of monthly and seasonal skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.,* **115,** 1825–1850.

——, M. E. Schlesinger, and X. Jiang, 1991: On greenhouse gas detection strategies. *Greenhouse-Gas-Induced-Climatic Change: A Critical Appraisal of Simulations and Observations,* M. E. Schleisinger, Ed., Elsevier Science, 537–558.

——, B. D. Santer, P. D. Jones, R. S. Bradley, and K. R. Briffa, 1996: Estimates of low frequency natural variability in near-surface air temperature. *Holocene,* **6,** 255–263.

——, G. Hegerl, T. Knutson, and S. Tett, 1999: Uncertainties levels in predicted patterns of anthropogenic climate change. *J. Geophys. Res.* (Atmos.)*,* submitted.

Bell, T. L., 1982: Optimal weighting of data to detect climate change: Application to the carbon dioxide problem. *J. Geophys. Res.,* **87,** 11 161–11 170.

——, 1986: Theory of optimal weighting of data to detect climate change. *J. Atmos. Sci.,* **43,** 1694–1710.

Briffa, K. R., P. D. Jones, F. H. Schweingruber, and T. J. Osborn, 1998: Influence of volcanic eruptions on Northern Hemisphere summer temperatures over the past 600 years. *Nature,* **393,** 450–455.

Christy, J. R., R. W. Spencer, and E. Lobl, 1998: Analysis of the merging procedure for the MSU daily temperature time series. *J. Climate,* **11,** 2016–2041.

Cubasch, U., and Coauthors, 1994: Monte Carlo climate forecasts with a global coupled ocean–atmosphere model. *Climate Dyn.,* **10,** 1–19.

——, G. C. Hegerl, and J. Waszkewitz, 1996: Prediction, detection, and regional assessment of anthropogenic climate change. *Geophysica,* **32,** 77–96.

——, R. Voss, J. Waszkewitz, and T. J. Crowley, 1997: Simulation of the influence of solar radiation variations on global climate with an ocean–atmosphere general circulation model. *Climate Dyn.,* **13,** 757–767.

Gaffen, D. J., M. A. Sargent, R. E. Habermann, and J. R. Lazante, 1999: Sensitivity of tropospheric and stratospheric trends to radiosonde data quality. *J. Climate,* in press.

Gates, W. L., J. S. Boyle, C. Covey, G. Dease, C. M. Doutriaux, R. S. Drach and others, 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP). *Bull. Amer. Meteor. Soc.,* **80,** 29–55.

Gibson, J. K., P. Kallberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serrano, 1997: ERA description. ECMWF Re-Analysis Project Rep. Series 1, 66 pp.

Hansen, J. E., and Coauthors, 1998: Climate forcings in the industrial era. *Proc. Natl. Acad. Sci.,* **95,** 12 753–12 758.

Hasselmann, K., 1979: On the signal-to-noise problem in atmospheric response studies. *Meteorology of Tropical Oceans,* D. B. Shaw, Ed., Royal Meteorological Society, 251–259.

——, 1993: Optimal fingerprints for the detection of time-dependent climate change. *J. Climate,* **6,** 1957–1969.

——, 1997: Multi-pattern fingerprint method for detection and attribution of climate change. *Climate Dyn.,* **13,** 601–611.
——, 1998: Conventional and Bayesian approach to climate change detection and attribution. *Quart. J. Roy. Meteor. Soc.,* **124,** 2541–2565.
——, and Coauthors, 1995: Detection of anthropogenic climate change using a fingerprint method. *Modern Dynamical Meteorology. Proceedings of the Wiin–Nielsen Symposium,* European Centre for Medium-Range Weather Forecasts, 203–221.
Hegerl, G. C., H. von Storch, K. Hasselmann, B. D. Santer, U. Cubasch, and P. D. Jones, 1996: Detecting anthropogenic climate change with an optimal fingerprint method. *J. Climate,* **9,** 2281–2306.
——, K. Hasselmann, U. Cubasch, J. F. B. Mitchell, E. Roeckner, R. Voss, and J. Waszkewitz, 1997: Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol, and solar forced climate change. *Climate Dyn.,* **13,** 613–634.
——, P. Stott, M. Allen, J. F. B. Mitchell, S. F. B. Tett, and U. Cubasch, 1999a: Detection and attribution of climate change: Sensitivity of results to climate model differences. *Climate Dyn.,* in press.
——, P. D. Jones, and T. P. Barnett, 1999b: Effect of sampling uncertainty on anthropogenic signal detection. *J. Climate,* submitted.
Houghton, J. T., G. J. Jenkins, and J. J. Ephraums, Eds., 1996: *Climate Change 1995: The IPCC Second Scientific Assessment.* Cambridge University Press, 572 pp.
Hoyt, D. V., and K. H. Shatten, 1993: A discussion of plausible solar irradiance variations. 1700–1992. *J. Geophys. Res.,* **98,** 18 895–18 906.
Johns, T. C., R. E. Carnell, J. F. Crossley, J. M. Gregory, J. F. B. Mitchell, C. A. Senior, S. F. B. Tett, and R. A. Wood, 1997: The second Hadley Centre coupled ocean–atmosphere GCM: Model descriptions, spinup and validation. *Climate Dyn.,* **13,** 103–134.
Jones, P. D., 1998: It was the best of times, it was the worst of times. *Science,* **280,** 544–545.
——, and G. C. Hegerl, 1998: Comparisons of two methods of removing anthropogenically related variability from the near-surface observational temperature field. *J. Geophys. Res.,* **103,** 13 777–13 786.
——, P. A. Groisman, M. Coughlan, N. Plummer, W.-C. Wang, and T. R. Karl, 1990: Assessment of urbanization effects in time series of surface air temperature over land. *Nature,* **347,** 434–439.
——, T. J. Osborn, and K. R. Briffa, 1997: Estimating sampling errors in large-scale temperature averages. *J. Climate,* **10,** 2548–2568.
——, K. R. Briffa, T. P. Barnett, and S. F. B. Tett, 1998: High-resolution palaeoclimatic records for the last millenium: Interpretation, integration and comparison with general circulation model control-run temperatures. *Holocene,* **8,** 455–471.
——, M. New, D. E. Parker, S. Martin, and I. G. Rigor, 1999: Surface air temperature and its changes over the past 150 years. *Rev. Geophys.,* **37,** 173–199.
Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.
Karl, T. R., R. W. Knight, and J. R. Christy, 1994: Global and hemispheric temperature trends—Uncertainties related to inadequate spatial sampling. *J. Climate,* **7,** 1144–1163.
Knutson, T. R., T. L. Delworth, K. Dixon, and R. J. Stouffer, 1999: Model assessment of regional surface trends (1949–1997). *J. Geophys. Res.* (Atmos.), in press.
Leroy, S., 1998: Detecting climate signals: Some Bayesian aspects. *J. Climate,* **11,** 640–651.
Manabe, S., and R. J. Stouffer, 1997: Climate variability of a coupled ocean–atmosphere–land surface model: Implication for the detection of global warming. *Bull. Amer. Meteor. Soc.,* **78,** 1177–1185.
Mann, M. E., R. S. Bradley, and M. K. Hughes, 1998: Global-scale patterns and climate forcing over the past six centuries. *Nature,* **392,** 779–787.
Meehl, G. A., G. J. Boer, C. Covey, M. Latif, and R. J. Stouffer, 1997: Intercomparison makes for a better climate model. *Eos, Trans. Amer. Geophys. Union,* **78,** 445–446, 451.
Mitchell, J. F. B., T. C. Johns, J. M. Gregory, and S. F. B. Tett, 1995: Climate response to increasing levels of greenhouse gases and sulfate aerosols. *Nature,* **376,** 501–504.
North, G. R., and K. Y. Kim, 1995: Detection of forced climate signals. Part II: Simulation results. *J. Climate,* **8,** 409–417.
——, and M. J. Stevens, 1998: Detecting climate signals in the surface temperature record. *J. Climate,* **11,** 563–577.
——, K. Y. Kim, S. P. Shen, and J. W. Hardin, 1995: Detection of forced climate signals. Part I: Filter theory. *J. Climate,* **8,** 401–408.
Parker, D. E., C. K. Folland, and M. Jackson, 1995: Marine surface temperature: Observed variations and data requirements. *Climate Change,* **31,** 559–600.
——, M. Gordon, D. P. N. Cullum, D. M. H. Sexton, C. K. Folland, and N. Rayner, 1997: A new global gridded radiosonde temperature database and recent temperature trends. *Geophys. Res. Lett.,* **24,** 1499–1502.
Roeckner, E., L. Bengtsson, J. Feichter, J. Lelievel, and H. Rodhe, 1999: Transient climate change simulations with a coupled atmosphere–ocean GCM including the tropospheric sulfur cycle. *J. Climate,* **12,** 3004–3032.
Santer, B. D., K. E. Taylor, T. M. L. Wigley, J. E. Penner, P. D. Jones, and U. Cubasch, 1995: Towards the detection and attribution of an anthropogenic effect on climate. *Climate Dyn.,* **12,** 77–100.
——, T. M. L. Wigley, T. M. L. Barnett, and E. Anyamba, 1996a: Detection of climate change and attribution of causes, *Climate Change 1995: The Science of Climate Change,* J. T. Houghton et al. Eds., Cambridge University Press, 407–444.
——, and Coauthors, 1996b: A search for human influences on the thermal structure in the atmosphere. *Nature,* **382,** 39–46.
——, J. J. Hnilo, T. M. L. Wigley, J. S. Boyle, C. Doutriaux, M. Fiorino, D. E. Parker, and K. E. Taylor, 1999: Uncertainties in observationally based estimates of temperature change in the free atmosphere. *J. Geophys. Res.,* **104** (D6), 6305–6333.
Sato, M., J. E. Hansen, M. P. McCormick, and J. Pollack, 1993: Stratospheric aerosol optical depths (1850–1990). *J. Geophys. Res.,* **98,** 22 987–22 994.
Stevens, M. J., and G. R. North, 1996: Detection of the climate response to the solar system. *J. Atmos. Sci.,* **53,** 2594–2608.
Stott, P. A., and S. F. B. Tett, 1998: Scale-dependent detection of climate change. *J. Climate,* **11,** 3282–3294.
——, ——, G. S. Jones, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell, 1999: Attribution of twentieth century temperature change to natural and anthropogenic causes. *Climate Dyn.,* submitted.

Stouffer, R. J., G. C. Hegerl, and S. F. B. Tett, 1999: A comparison of surface air temperature variability in three 1000-year coupled ocean–atmosphere model integrations. *J. Climate,* in press.

Tett, S. F. B., J. F. B. Mitchell, D. E. Parker, and M. R. Allen, 1996: Human influence on the atmospheric vertical temperature structure: Detection and observations. *Science,* **274,** 1170–1173.

——, P. A. Stott, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell, 1999: Causes of twentieth century temperature change near the earth's surface. *Nature,* **399,** 569–572.

Voss, R., R. Sausen, and U. Cubasch, 1997: Periodically synchronously coupled integrations with the atmosphere–ocean general circulation model ECHAM3/LSG. *Climate Dyn.,* **14,** 249–266.

Wigley, T. M. L., and T. P. Barnett, 1990: Detection of the greenhouse effect in the observations. *Scientific Assessment of Climate Change: Intergovernmental Panel on Climate Change,* IPCC Working Group I, WMO/UNEP, 239–255.

——, P. J. Jaumann, B. D. Santer, and K. E. Taylor, 1998a: Relative detectability of greenhouse-gas and aerosol climate change signals. *Climate Dyn.,* **14,** 781–790.

——, R. L. Smith, and B. D. Santer, 1998b: Anthropogenic influence on the autocorrelation structure of hemispheric-mean temperatures. *Science,* **282,** 1676–1679.