Detecting Greenhouse-Gas-Induced Climate Change with an Optimal Fingerprint Method

GABRIELE C. HEGERL,* HANS VON STORCH,* KLAUS HASSELMANN,*
BENJAMIN D. SANTER,+ ULRICH CUBASCH,& AND PHILIP D. JONES@

*Max-Planck-Institut für Meteorologie, Hamburg, Germany.
+PCMDI/Lawrence Livermore National Laboratory, Livermore, California.
&Deutsches Klimarechenzentrum, Hamburg, Germany.
@Climatic Research Unit, University of East Anglia, Norwich, United Kingdom.

(Manuscript received 26 August 1994, in final form 20 March 1996)

ABSTRACT

A strategy using statistically optimal fingerprints to detect anthropogenic climate change is outlined and applied to near-surface temperature trends. The components of this strategy include observations, information about natural climate variability, and a "guess pattern" representing the expected time–space pattern of anthropogenic climate change. The expected anthropogenic climate change is identified through projection of the observations onto an appropriate optimal fingerprint, yielding a scalar-detection variable. The statistically optimal fingerprint is obtained by weighting the components of the guess pattern (truncated to some small-dimensional space) toward low-noise directions. The null hypothesis that the observed climate change is part of natural climate variability is then tested.

This strategy is applied to detecting a greenhouse-gas-induced climate change in the spatial pattern of near-surface temperature trends defined for time intervals of 15–30 years. The expected pattern of climate change is derived from a transient simulation with a coupled ocean–atmosphere general circulation model. Global gridded near-surface temperature observations are used to represent the observed climate change. Information on the natural variability needed to establish the statistics of the detection variable is extracted from long control simulations of coupled ocean–atmosphere models and, additionally, from the observations themselves (from which an estimated greenhouse warming signal has been removed). While the model control simulations contain only variability caused by the internal dynamics of the atmosphere–ocean system, the observations additionally contain the response to various external forcings (e.g., volcanic eruptions, changes in solar radiation, and residual anthropogenic forcing). The resulting estimate of climate noise has large uncertainties but is qualitatively the best the authors can presently offer.

The null hypothesis that the latest observed 20-yr and 30-yr trend of near-surface temperature (ending in 1994) is part of natural variability is rejected with a risk of less than 2.5% to 5% (the 5% level is derived from the variability of one model control simulation dominated by a questionable extreme event). In other words, the probability that the warming is due to our estimated natural variability is less than 2.5% to 5%. The increase in the signal-to-noise ratio by optimization of the fingerprint is of the order of 10%–30% in most cases.

The predicted signals are dominated by the global mean component; the pattern correlation excluding the global mean is positive but not very high. Both the evolution of the detection variable and also the pattern correlation results are consistent with the model prediction for greenhouse-gas-induced climate change. However, in order to attribute the observed warming uniquely to anthropogenic greenhouse gas forcing, more information on the climate's response to other forcing mechanisms (e.g., changes in solar radiation, volcanic, or anthropogenic sulfate aerosols) and their interaction is needed.

It is concluded that a statistically significant externally induced warming has been observed, but our caveat that the estimate of the internal climate variability is still uncertain is emphasized.

1. Introduction

Numerous model simulations predict a human-induced change in the mean state of climate due to the increasing concentration of greenhouse gases in the atmosphere. Many previous studies have tried to detect such a change in observations, usually in global, annually averaged near-surface temperature data. The observed 100-yr trend in global mean temperature has been shown to be larger than can be explained by the noise generated by a simple climate model (Wigley and Raper 1990, 1991) or by a coupled ocean–atmosphere general circulation model (CGCM; Stouffer et al. 1994), but it has not been possible to attribute the change to the greenhouse gas forcing. Rather than preselecting some arbitrary climate index (e.g., global mean temperature) for detection, several authors have proposed the use of fingerprint methods, in which our knowledge of the expected pattern of climate change

Corresponding author address: Dr. Gabriele C. Hegerl, Max-Planck-Institut für Meteorologie, Bundesstraße 55, D-20146 Hamburg, Germany.

is used to enhance the possibility of detecting and attributing anthropogenic climate change (e.g., Barnett 1986, 1991; Barnett and Schlesinger 1987; Hasselmann 1979, 1993; Madden and Ramanathan 1980; Barnett et al. 1991; Santer et al. 1993b). In a summary of the state of the art of anthropogenic climate change detection, Wigley and Barnett (1990) concluded that at that time the detection efforts were not yet statistically convincing. The basic difficulties and uncertainties that impede the detection of climate change are discussed in Santer et al. (1993a). In two parallel papers, Santer et al. (1993b, 1995a) tried to detect anthropogenic climate change using centered and uncentered pattern correlations between the time-varying spatial patterns of observed near-surface temperature changes and the time-independent signal patterns derived from a number of equilibrium experiments and one transient response experiment using atmospheric general circulation models (AGCMs) coupled to a simple ocean. Here "uncentered correlation" refers to a correlation of the pattern including the spatial mean (sometimes called "pattern similarity index," Barnett et al. 1991); "centered correlation" refers to the correlation of deviation patterns, where the spatial mean has been subtracted. While the earlier studies considered only signals resulting from a doubling of atmospheric CO_2, more recent work has also analyzed experiments with combined CO_2–sulfate aerosol forcing. If the spatial means were subtracted, little evidence for time-increasing spatial congruence between the simulated annual mean greenhouse warming patterns and observed near-surface temperature changes was found in either of the two studies. However, in the case of the combined CO_2–sulfate aerosol forcing, the seasonal temperature change patterns yielded multidecadal positive trends in the pattern correlation statistics (Santer et al. 1995a). Karoly et al. (1994) also found positive trends in the pattern correlation between observed and simulated annual and zonal mean vertical profiles of atmospheric temperatures for equilibrium greenhouse warming experiments. Santer et al. (1996b) found a similar agreement using combined forcing simulations. A full summary of the findings of recent pattern-based detection studies is given in the 1995 IPCC report (Houghton and Meira Filho 1996; cf. Santer et al. 1996a).

Simulations with more realistic CGCMs have greatly improved the two basic inputs needed for signal detection: they have provided estimates of the time-dependent spatial pattern of the climate response under transient global warming conditions and the space–time structure of the natural variability of the climate system. With this new information, attempting a quantitative statistical detection of greenhouse-gas-induced global warming using advanced fingerprint techniques has gained in feasibility.

We propose and apply in the following a method for detecting anthropogenic climate change in accordance with a general strategy described in Pennell et al. (1993). The method uses an optimal fingerprint approach introduced by Hasselmann (1979). In a later paper, Hasselmann (1993) extended the optimization formalism for spatially dependent signal patterns to the general space–time-dependent case and to the simultaneous application of several fingerprints (e.g., to superimposed signals generated by the individual forcings of different greenhouse gases and sulfate aerosols). Similar methods have been proposed by Bell (1982, 1986) and North et al. (1995). The optimal fingerprint strategy has been previously applied to ocean climate change indices derived from CGCM global warming simulations using natural variability estimates from ocean GCM simulations (Santer et al. 1995b). The authors found that although the method is rather sensitive to the noise characteristics used for the optimization, the multivariate optimal fingerprint approach yielded a marked detection enhancement in the model world.

We apply the optimal fingerprint here to near-surface temperature observations. CGCM simulations predict accelerating trends of global mean temperature in the second half of this century due to anthropogenic greenhouse gas forcing (e.g., Cubasch et al. 1992, 1994, 1995) and, more recently, greenhouse gas and aerosol forcing (Mitchell et al. 1995; Hasselmann et al. 1995). Thus, rather than considering the 100-yr trend, as most earlier investigations, we focus on the temperature trends for time intervals of 15–30 years, which emphasize the recent warming signal. The chosen trend lengths are a compromise between an emphasis on the accelerating nature of the anthropogenic signal (requiring a short time interval) and the reduction of natural variability noise (requiring a large time interval). We attempt to identify a model-derived greenhouse-warming-only signal in the observed data; the extension to a signal including both greenhouse warming and aerosol cooling is straightforward (Hasselmann et al. 1995). The greenhouse gas forcing is represented in the model simulation in terms of "equivalent CO_2" concentrations (Houghton et al. 1990), in which the net radiative forcing contribution of all anthropogenic greenhouse gases are expressed in terms of an increased CO_2 concentration.

All a priori information regarding the anticipated greenhouse warming signal is derived from model simulations. Special efforts are devoted to estimating natural variability since the outcome of any statistical significance test is crucially dependent on the climate noise estimate (see, e.g., Santer et al. 1995b). Estimates of the natural variability of the climate system are gained from both model and observational data. In this paper, we refer to unforced variability associated with the internal dynamics of the ocean–atmosphere system as "natural variability." We are aware that, ideally, "natural variability" would refer to all naturally occurring climate fluctuations, also including naturally occurring external forcings like volcanic eruptions or solar radiation changes. However, at the present time,

these have not yet been incorporated into long climate model simulations. Uncertainties exist whether present climate models are able to reliably simulate internal climate variability on timescales of decades to centuries and to the extent to which observations are influenced by various possible external forcing mechanisms. The observations since the last century will presumably contain an evolving greenhouse gas signal and additionally the signatures of other anthropogenic forcings, for example sulfate aerosols. We subtract an estimate of the greenhouse gas signal from the observed data using a simple response model that has been fitted to a transient climate change simulation (Hasselmann et al. 1993; Cubasch et al. 1995; Tahvonen et al. 1993).

The structure of this paper is as follows. Section 2 describes the strategy for detecting anthropogenic climate change. In section 3, we apply this strategy to near-surface temperature data using as a fingerprint the spatial pattern of expected trends. In section 4, an optimal fingerprint is applied and the results are discussed, from which we infer that a statistically significant warming has occurred. Section 5 addresses the question whether the warming can be attributed to anthropogenic greenhouse gas forcing. Our conclusions are summarized in section 6.

2. Detection strategy

The first steps of the detection strategy are to choose variables that permit a distinction between the signal of anthropogenic climate change and the climate's natural variability (section 2a) and define a "guess pattern" representing the anticipated climate change signal in terms of these variables (section 2b). The fingerprint (which is a contravariant vector contrary to the covariant guess pattern and signal pattern; Hasselmann 1979) is then either chosen as identical to the guess pattern ("guess-pattern fingerprint") or statistically optimized ("optimal fingerprint," section 2c). Finally, we test the null hypothesis that the observed climate change originates from natural variability (section 2d).

a. *Selection of climate change variables*

Formally, all climate variables for which sufficiently reliable observations and simulations exist are acceptable for the fingerprint method. An overview of the quality criteria for both observed and model data are given by, for example, Wigley and Barnett (1990) and Santer et al. (1991, 1993a), while investigations of suitable variables for detection purposes are given in Barnett et al. (1991) and Santer et al. (1994). If a large set of variables is used, the optimal fingerprint method automatically assigns higher weight to variables with high signal to noise ratio. However, to simplify the analysis, it is advisable in practice to reduce the number of degrees of freedom by preselecting the variables on the basis of prior estimates of signal to noise ratios.

We use only one variable, near-surface temperature, for the detection test. Near-surface temperature has been observed for a comparatively long time, providing reasonably good information on the time-dependence of the observed climate change and on observed climate variability on decadal timescales. Also, results from a control run and a time-dependent greenhouse warming experiment using the Hamburg coupled ocean–atmosphere model (Cubasch et al., 1992) suggest that the signal to noise ratio is higher for this variable than for other variables such as sea level pressure (Santer et al. 1994).

We consider then a climate vector $\Psi = (\Psi_1 \cdots \Psi_n)$ in the "detection space" V. In the climate change detection method of Hasselmann (1993), the climate change signal is regarded as a space–time-dependent pattern, consisting of the values of the selected variables, in this case the $(2 + 1)$ dimensional near-surface temperature field $T(\mathbf{x}, t)$. The spatial representation could be, for example, a (gridded) or irregular (stations) distribution or an expansion with respect to spherical harmonics or EOFs (empirical orthogonal functions); while the time-behavior of the variable could be represented by average values over certain discrete time intervals (for example annual mean temperatures) or by Fourier coefficients.

However, in the present paper we apply only a simplified version of Hasselmann's technique in which the time dependence of the signal is represented locally as a linear trend over some suitably chosen, relatively short time interval τ. If τ is much smaller than the observation period, we can obtain an estimate of the natural climate variability also from the observations. We consider this an advantage that is worthwhile foregoing the possibility of optimizing the fingerprint with respect to the time-integration (cf. Hasselmann 1993). The choice of the trend length of 15–30 years is a compromise between obtaining a reasonable natural variability estimate from the relatively short ($\sim$130 years) observation period on the one hand and the stronger noise component in short trends on the other.

In practice, the original data must be truncated also in space to some smaller dimensional space to apply the optimal fingerprint method (Hasselmann 1979, 1993). We come back to this point below.

b. *The fingerprint approach*

Fingerprint methods use some univariate pattern similarity index, or, more generally, a low-dimensional detection vector to compare the observations with the expected climate change pattern (see, e.g., Hasselmann 1979, 1993; Barnett et al. 1991). This operation can also be understood as applying a filter to the observations, the optimal fingerprint representing an optimally matched filter (Hasselmann 1979; North et al. 1995). This has advantages compared with the two extremes of trying to assess a significant climate change in the

full climate variable space or simply using a mean value.

- In general, the power of statistical significance tests in the original selected n-dimensional detection space V is very low. The significance of a given but unknown signal embedded in a noise background decreases rapidly with increasing dimension of the space (Hasselmann 1979; von Storch and Roeckner 1983; Bell 1986). Fingerprint approaches reduce the detection problem to a univariate or low-dimensional problem in the detection variable.
- Contrary to approaches using, for example, a mean value only, fingerprint methods use the full expected pattern of climate change and thus increase the chances of detecting an anthropogenic climate change and attributing it to the assumed forcing.

Hasselmann proposed projecting the n-dimensional space V onto a smaller number of variables by forming the scalar products $d_p = \mathbf{f}_p^{\mathrm{T}}\boldsymbol{\Psi}$ of the vector $\boldsymbol{\Psi}$ in V with a relatively small number of suitably selected fingerprints $\mathbf{f}_p$, $p = 1, 2, \cdots$. In this paper, however, we shall consider only a single fingerprint $\mathbf{f}$ [the general multifingerprint case is discussed in Hasselmann (1979) and, in the present terminology, in Hasselmann (1993)]. This yields the single scalar-detection variable:

$$d = \mathbf{f}^{\mathrm{T}}\boldsymbol{\Psi}. \tag{1}$$

We will derive the optimal fingerprint pattern from an uncentered signal pattern, that is, without subtraction of the spatial mean. Thus, the detection variable d (for a nonoptimized fingerprint) is similar to the $C(t)$ statistic used in Barnett et al. (1991) and in Santer et al. (1993b, 1995a). As pointed out by Santer et al. (1993b) and Barnett et al. (1991), such a statistic will reflect the increase in global mean temperature in the observations as well as the pattern correlation. The contribution of the pattern mean to the detection variable depends on the structure of the fingerprint and can range from a 100% (for a spatially uniform fingerprint) to zero (for a fingerprint with zero mean). In the case of the greenhouse warming signal, the mean warming is expected to be the dominant component of the greenhouse-gas-induced climate change (North and Kim 1995). We take the point of view here that both the global mean and the spatial structure of the deviation about the mean contribute to the description of the greenhouse warming signal and that using both increases the chances of detecting climate change. We discuss the contribution of each component to our results, as well as the implication for attributing a significant change to the greenhouse gas forcing later in the paper.

From a statistical point of view, the fingerprint is merely a tool. It is important only that it is defined a priori, without reference to the observed data, in order to ensure a statistically independent detection test. A fingerprint cannot be statistically "right" or "wrong" but only "well chosen" or "inadequate." A good choice of the fingerprint, based on the best available knowledge of the selected variable space, increases the chances for detection. We have chosen the fingerprint from of the output of a greenhouse warming simulation. It is important to note that if the observed data contain responses in addition to the greenhouse warming signal, for example, an anthropogenic aerosol signal, a fingerprint that is based only on the anticipated greenhouse warming signal is not "wrong" but merely suboptimal.

c. *The optimal fingerprint*

If the fingerprint shares some common features with typical patterns of natural climate variability, the detection variable will be contaminated by natural variability noise. A statistically optimal fingerprint, which maximizes the signal to noise ratio, can be obtained by rotating the fingerprint relative to the anticipated signal pattern away from the directions of high noise.

Hasselmann (1979, 1993) showed that the square signal to noise ratio

$$R^2 = \frac{d_S^2}{\epsilon(\tilde{d}^2)} \tag{2}$$

is maximized by choosing

$$\mathbf{f} = \mathbf{C}^{-1}\mathbf{g} \quad \text{where} \quad \mathbf{C} = \epsilon(\tilde{\boldsymbol{\Psi}}\tilde{\boldsymbol{\Psi}}^{\mathrm{T}}). \tag{3}$$

Here $\mathbf{g}$ denotes the "guess pattern," which is assumed to represent the expected signal of climate change $\boldsymbol{\Psi}_S$ in V (possibly up to an unknown amplitude); $\mathbf{C}$ is the covariance matrix of the random climate noise variable $\tilde{\boldsymbol{\Psi}}$; $\tilde{d} = \mathbf{f}^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}$, $d = \mathbf{f}^{\mathrm{T}}\boldsymbol{\Psi}$, and $d_S = \mathbf{f}^{\mathrm{T}}\boldsymbol{\Psi}_S$ are the detection variable for climate noise, the observed climate change, and the pure signal of climate change, respectively; and ϵ denotes the expectation. The approach has been applied by Hannoschöck and Frankignoul (1985) to AGCM sea surface temperature response experiments and by Santer et al. (1995b) to the detection of ocean global warming in a model simulation study.

The relation (2) is more easily interpreted if $\tilde{\boldsymbol{\Psi}}$ is represented in its EOF coordinates, so that the matrices $\mathbf{C}$ and $\mathbf{C}^{-1}$ are diagonal. The optimal fingerprint is then obtained by weighting each coordinate of the guess-pattern fingerprint g_i by the inverse of the associated EOF-eigenvalue (variance) λ_i: $f_i = g_i/\lambda_i$, thereby enhancing components with small noise λ_i.

The fingerprint is optimal only if the guess-pattern fingerprint does indeed represent the correct signal of anthropogenic climate change. An error in the guess pattern, for example, due to noise in estimating the pattern from a CGCM simulation, results in a suboptimal fingerprint and thus in an underestimate of the true sig-

nal to noise ratio. In this sense the detection method is conservative.

We normalize to unity both the guess-pattern fingerprint and the optimal fingerprint before computing the detection variable. Then signal and noise can be compared prior to and after optimization by means of the detection variable. In both cases, the value of the detection variable can be used for estimating the amplitude of the climate change signal (Hasselmann 1993; North et al. 1995; Hegerl and North 1996, manuscript submitted to *J. Climate*). Note that if the optimal fingerprint is applied, the value of the detection variable will decrease for the climate change signal. The benefit of rotation is that the noise amplitude will decrease stronger than the signal amplitude.

In practice, the covariance matrix $\mathbf{C}$ is unknown and has to be estimated from observed and model data. This introduces the following problems if the number of independent realizations of natural variability is small:

- The variance of high-indexed EOFs will be underestimated (North et al. 1982; von Storch and Hannoschöck 1985), yielding unrealistically high fingerprint components for these EOFs. The dimension of the covariance matrix $\mathbf{C}$ must therefore be kept small by suitably truncating the analysis space. Of course, the reduced phase space must still be able to represent most of the anticipated signal. This is the case for the space spanned by the dominant EOFs of the signal (Santer et al. 1994) but not necessarily for the EOFs of natural variability.
- The problem that the optimal fingerprint may be rotated into a direction for which the natural variability is poorly sampled, rather than genuinely low, can be exacerbated if care is not taken to use statistically independent data. If the same variability information is used for optimizing the fingerprint and for computing the statistics of the detection variable, the signal to noise estimate will be dominated by the emphasis of the optimal fingerprint on the poorly sampled low-variability components. This may lead to a severe overestimate of the signal to noise ratio (Hannoschöck and Frankignoul 1985; Bell 1986; Hasselmann 1993), which is similar that the bias caused by using data to fit a (e.g., statistical) model and then testing the model with the same data. We therefore chose independent data for estimating the covariance matrix and for the statistical test.

Having chosen the optimal fingerprint, we can now compute the components needed for the statistical detection test.

- The detection variable for the observed warming, that is, for that part of the observed data that we expect to be most representative of the anticipated anthropogenic climate change (e.g., in our application, the latest observed temperature trend) is computed.
- Samples of the detection variable representing the natural variability of the detection variable in the absence of external forcing are also computed. To obtain these, we use both observations and the output of several "control" simulations without external forcing, which are sufficiently long to resolve variability on all relevant timescales. Present CGCMs use a rather coarse grid for century-timescale simulations, which is inadequate for the description of the small-space-scale, short-timescale features of natural variability. However, it is generally believed that models reproduce the space–time statistics of natural variability on large space and long time scales (months to years) reasonably realistic (Gates et al. 1992, 1993). The verification of variability of CGCMs on decadal to century timescales is more difficult since the instrumental record is relatively short, while paleoclimatic data are sparse and often of limited quality (see, e.g., Santer et al. 1993a; Barnett et al. 1996).

In estimating the natural variability from observed data, we face the difficulty that the data also contains responses to diverse external forcing mechanisms. However, in some cases the forcing and the climate response can be estimated, and if the response is assumed to be linearly superimposed on the data, an estimate of the response can be subtracted. An example for such a procedure is given in section 3.

d. Statistical test

The information gained in the previous steps of the procedure is now combined in a statistical test. We test the null hypothesis that the observed climate change originates from natural variability. We apply a parametric statistic (Pennell et al. 1993), that is, we use a statistical model—the Gaussian distribution—that is appropriate for many climatological applications. If the multivariate climate vector in the detection space V is Gaussian, so is the linearly derived detection variable $\tilde{d} = \mathbf{f}^{\mathrm{T}}\tilde{\mathbf{\Psi}}$, where $\tilde{\mathbf{\Psi}}$ is a realization of natural variability. In fact, since $\tilde{d}$ represents an averaging process, the Central Limit Theorem suggests that the Gaussian assumption will probably be a better approximation for $\tilde{d}$ than for $\tilde{\mathbf{\Psi}}$.

If the mean and the standard deviation σ are known, we can choose a risk $1 - p$, and calculate the smallest interval Ω_p^N for which

$$\mathrm{prob}(\tilde{d} \in \Omega_p^N) = p. \tag{4}$$

The null-hypothesis is then rejected if the detection variable for the observed climate change, $d = \mathbf{f}^{\mathrm{T}}\mathbf{\Psi}$, is not contained in Ω_p^N. Since a positive value of the detection variable (1) is expected if greenhouse warming is occurring, a one-tailed test is performed. For example, for Gaussian statistics the null hypothesis is rejected with a risk of 2.5%, if $d > 1.96\sigma$. This is the risk level we choose for rejecting the null hypothesis in this paper.

If the parameters of the Gaussian distribution are estimated by a finite and independent set of samples of natural variability $\tilde{d}_i = \mathbf{f}^T\tilde{\mathbf{\Psi}}_i$, $i = 1, \cdots \nu$, a t test with $(\nu - 1)$ degrees of freedom can be performed. If the samples are interdependent (i.e., autocorrelated) the concept of the t test breaks down (Thiebaux and Zwiers 1984; Zwiers and von Storch 1995). In this case the statistics of the detection variable can be estimated by Monte Carlo simulations (see section 4b). Since both assumptions—that the statistic is Gaussian and that the interdependence of the data can be modeled by an AR(1) process—are hard to verify in practice, we also compare the results of our detection approach to those obtained if the sampling distributions were used.

If the null hypothesis cannot be rejected, we conclude that a statistically significant global warming cannot be detected at the chosen confidence (risk) level. If the null hypothesis is rejected, we have shown only that the observed representation of climate change deviates significantly from (our estimate of) natural climate variability. In section 5 we discuss how this observed climate change could be uniquely attributed to a particular anthropogenic forcing, for example, to greenhouse gas forcing. This requires detailed investigations of the other competing mechanisms, which is not attempted in the present paper.

3. Application to near-surface temperature

In the following, we apply the detection strategy to trends of near-surface temperature. Following the approach outlined in the last section, we first discuss the observed and model data and their time–space representation (sections 3a,b). Then we choose the guess pattern (section 3c) and attempt to assess the natural climate variability (section 3d). In section 3e, a suitable small subspace is defined for computing the optimal fingerprint.

a. Model and observed data

As observed data, we use the global gridded anomalies of monthly mean near-surface temperature data compiled by Jones (Jones et al. 1991; Jones and Briffa 1992; Briffa and Jones 1993; Jones 1994a,b). The data are in the form of anomalies with respect to the mean for the years 1950–1979. The observations have been assessed for both random and systematic errors (cf. Jones et al. 1986a,b; 1991) and are generally considered to be reliable. Global-mean near-surface temperature curves and the spatial pattern of temperature changes over the last decade based on these data are presented in the IPCC Scientific Assessment (Folland et al. 1992).

The data are available on a 5° × 5° global grid; gridpoints with inadequate data coverage are indicated as missing. We use the grid of the observed data as a common grid for both observed and model data (i.e., the model data are transformed to the 5° × 5° grid). The data range from the year 1854 to 1994, with data coverage changing in time. We analyze the data both with respect to climate change and to assess the natural variability.

As model data, we have used the output of three different CGCMs.

- The model ECHAM/LSG (in the following referred to as HAML) consists of the ECHAM atmospheric general circulation model, cycle 1 (Roeckner et al. 1992), and the ocean model LSG ("Large Scale Geostrophic," Maier-Reimer et al. 1993). ECHAM is a spectral T21 model (with associated 5.6° Gaussian grid) with 19 levels in the vertical. The LSG ocean model has 11 layers in the vertical and a horizontal resolution of 5° × 5°. It includes an ice model with simple rheology. Flux corrections are included in the ocean–atmosphere coupling to avoid model drift (Cubasch et al. 1992). The ECHAM/LSG model has been used in a number of climate change experiments (Cubasch et al. 1992, 1994, 1995) and in a multicentury control integration with an equivalent atmospheric CO_2 concentration fixed at the present day level (von Storch 1994).
- The model ECHAM2/OPYC (Lunkeit et al. 1995, hereafter HAMO) features an updated version of the ECHAM–atmosphere model (ECHAM2) and an ocean model in isopycnic coordinates (OPYC, Oberhuber 1993a,b). The spatial resolution of the atmospheric model is the same as in the HAML model. The ocean model has a vertical resolution of nine layers and an explicit mixed layer; its horizontal resolution is twice that of the atmosphere model, increasing toward the equator. A Hibler-type ice model is included. Flux corrections are applied as in the HAML. The model has been used in a 210-yr control simulation and a climate change experiment (Lunkeit et al. 1995).
- The GFDL-coupled CGCM consists of a spectral AGCM on a 7.5° longitude × 4.5° latitude Gaussian grid coupled to a primitive equation OGCM of a 3.75° longitude × 4.5° latitude resolution. The models are coupled using flux correction. The model and its variability are described in Stouffer et al. (1994), Manabe and Stouffer (1996), and Delworth et al. (1993).

Our anticipated anthropogenic climate change signal is taken as the output of a climate change simulation with the HAML model (Cubasch et al. 1995). The experiment, referred to as "EIN" (Early Industrialization), starts from the equivalent atmospheric CO_2 concentration for 1935 and uses the observed changes in equivalent CO_2 concentration from 1935 to 1985 and, for the remaining 100 years (i.e., for the model years 1986–2085), the greenhouse gas forcing of the IPCC Scenario A ("business as usual," Houghton et al. 1990).

To estimate the internal variability of climate we use as model data two 1000-yr control simulations with the

HAML and GFDL models and the 210-yr control simulation with HAMO. All control simulations exhibit complex variability on a wide range of space and timescales. A particular feature of HAMO is the strong contribution to the total variance from a single event: a strong global average temperature decrease of roughly 0.4°C in the eleventh decade (see Fig. 1b). This is caused by an inflow of fresh Arctic surface water into the North Atlantic, resulting in a decline in the thermohaline circulation. It could conceivably be related to observed events such as the Great Salinity Anomaly described by Dickson et al. (1988)—although this involved a shorter timescale (10–15 years) and a smaller amplitude of the temperature anomaly—or to a collapse of the North Atlantic thermohaline circulation, as produced in simulations of the Younger Dryas event (Maier-Reimer and Mikolajewicz 1989; Manabe and Stouffer 1988). A discussion of the HAMO event is given in Lunkeit et al. (1995). In the context of the present statistical evaluation, it is regarded as an extreme event of internal model variability, even though the event is of questionable realism.

b. Representation of the climate change signal

Transient greenhouse gas simulations predict an accelerating greenhouse gas signal (Cubasch et al. 1992, 1994, 1995). Thus, we expect the temperature trends to evolve from small trends in early industrial times to increasingly stronger trends as we progress from the present into the future. This is caused by a combination of the increase of the forcing and the delayed response of the climate system (see section 3d).

We apply the detection strategy to 15-yr, 20-yr, and 30-yr linear trends. Rather than longer trends, those relatively short trends emphasize the expected increasing warming signal and enable us additionally to derive an estimate of natural climate variability from the observations (section 2a). If the trend length τ is smaller than the total length of the climate change simulation, we can regard the detection variable d as a slowly varying function of time. This has the advantage that we can consider the gradual increase of the significance of d with time.

The trends are calculated from annual mean values. A linear trend is fit by least squares (regression line) to each grid point and for each interval of the respective trend length. For each computed trend we require that the data gaps are at most one year (this implies, for example, that at least eight individual years of data are needed to compute a 15-yr trend) and that there is at least one seasonal mean (i.e., at least one month) available per year, a procedure similar to that used by Parker et al. (1994). The results were insensitive to using more stringent criteria.

Grid points where observed trends are missing after 1949 have been excluded from the detection space. This is a compromise between the following two conflicting requirements.

- Nonrandom changes in data density as a function of time, that is, the systematic absence of observations in coherent areas of the globe for parts of the observation period, can cause errors in the estimate of the statistics of the detection variable. An example of such an effect will be given in section 4. Thus, we need a natural variability dataset that is as complete a representation as possible of the detection space.
- We wish to represent the distinctive features of the anticipated greenhouse gas signal, per example, the land–sea contrast and the cooling in the North Atlantic. If we accept, for example, only data in the detection space that have been continually observed since 1910, the resulting data-covered area will be quite small. By restricting the requirement of continuous coverage of the detection space to the time since 1949, we miss mainly observations in the high southern latitudes, which have been regularly observed only recently and where the model data are also less likely to be reliable.

We denote the resulting detection space of all grid points with continuous observed trends since 1949 V'. This yields a description of temperature on a planetary scale but excluding most Antarctic regions, parts of the Arctic, parts of the Pacific, and few smaller areas in the interior of Africa and South America (see Fig. 3): it covers approximately 75% of the globe. Using slightly different data coverage (e.g., by demanding continuous coverage a decade earlier or later) yielded only slightly modified results. The data have been weighted in proportion to the areal coverage that they represent. Figure 1 shows the mean near surface temperature of V' for the observations and for the three control simulations.

We anticipate that the latest observed 15-yr, 20-yr and 30-yr trends, ending with the year 1994, contain contributions from greenhouse-gas-induced climate change. Figure 2a shows, for example, the latest observed spatial pattern of 30-yr trends. Figure 2b shows, in contrast, the observed 30-yr trend pattern for the earlier time interval 1916–1945. This trend pattern, although also characterized by a high mean temperature increase, has a different spatial signature, with strongest warming in the northern high latitudes. We shall discuss the impact of these differences in the warming patterns on the detection variable in section 4.

c. Choice of the guess pattern

We define our anticipated pattern of anthropogenic greenhouse warming as the temperature change between the decadal mean temperatures from 1986 to 1995 and the decade 2076 to 2085 for the EIN simulation (Fig. 3). The pattern is very similar to the first EOF of the EIN simulation, which (contrary to the second and higher EOFs) explains the major proportion of the variance of the simulation (90% of the last 100 years). Its principal component provides a description of the warming trend in the model (Cubasch et al.

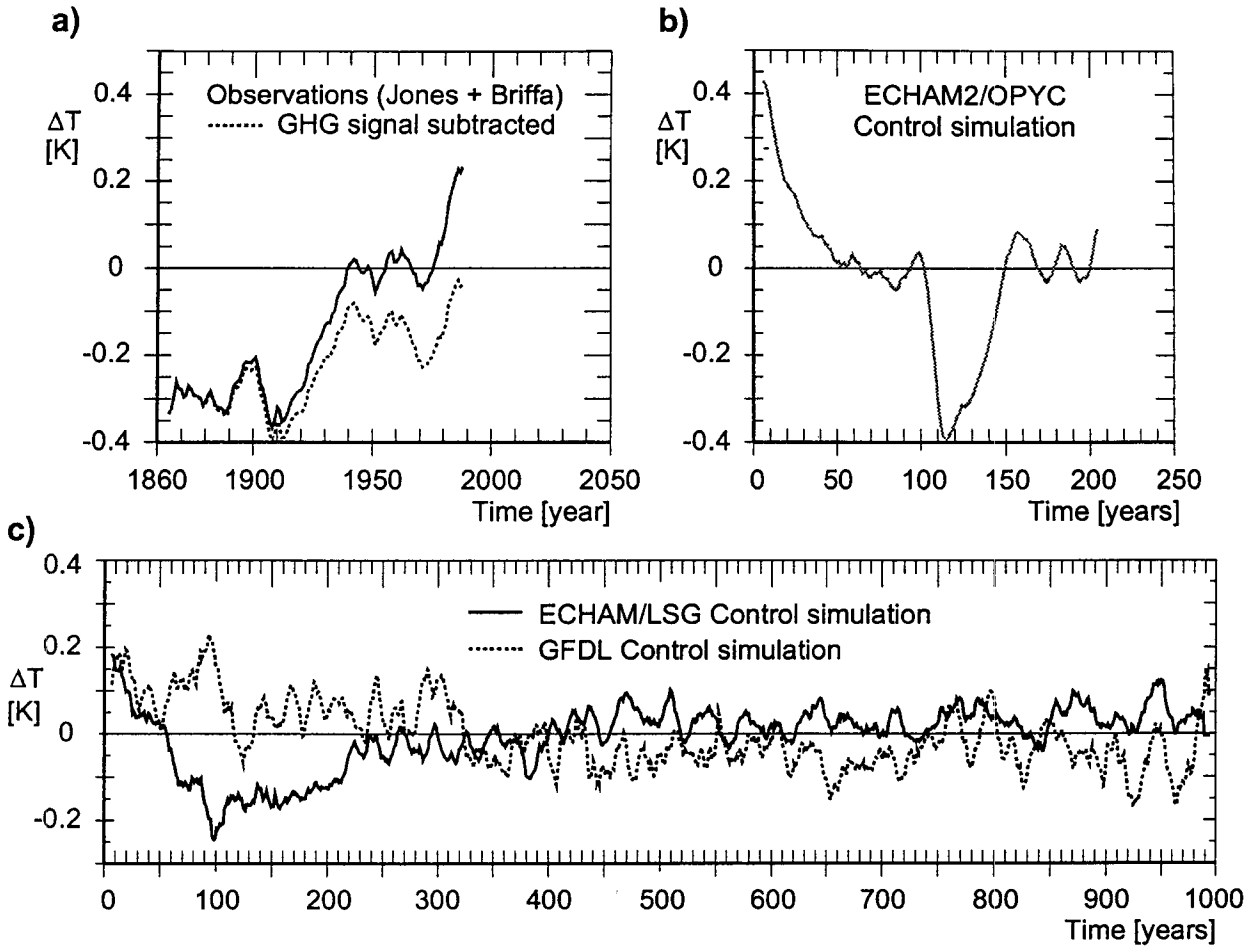

FIG. 1. Time evolution of global mean near-surface temperature for the observations (a) for the ECHAM2/OPYC (HAMO) control simulation (b) and for two 1000-yr control simulations with the GFDL and ECHAM/LSG (HAML) model (c), smoothed by an 11-yr running mean. Note the different scales on the time axis. The dashed line in (a) shows the observed global mean temperature after the greenhouse gas signal has been subtracted (cf. section 3d). All global mean temperatures are calculated using only grid points for which reliable observed trends could be calculated since 1949 (cf. Fig. 3).

1995). Thus, our choice of the guess pattern should yield a representative pattern for the temperature trends expected in the context of greenhouse warming. The use of a guess-pattern calculated from short trend patterns of the EIN simulation is not advisable since short trends from a single simulation are strongly contaminated by the overlying climate noise. Theoretical considerations also show that such a noise contamination of the guess pattern leads to deteriorated results (section 2c; Hegerl and North 1996, manuscript submitted to *J. Climate*).

We chose the EIN model simulation rather than earlier simulations with the Hamburg CGCM, such as the scenario A simulation of Cubasch et al. (1992), because the EIN simulation had the more clearly developed greenhouse gas signal (Cubasch et al. 1995). The pattern of temperature change is nevertheless very similar in both simulations (centered spatial pattern correlation of 0.8 globally, differences mainly in high southern latitudes not represented in V'), and the detection results presented here are essentially identical to the results using a fingerprint derived from the earlier scenario A simulation.

The warming patterns for both HAML simulations are also similar to the patterns of anthropogenic greenhouse warming simulations with other CGCMs. Table 1 shows the (centered and uncentered) pattern correlations for summer and winter temperature changes between the earlier scenario A simulation and those calculated with various other recent CGCMs. The models used for these simulations are described in Gates et al. (1992, 1993). All models agree fairly well in winter. The lower correlations in summer result from the different behavior of the models in high latitudes and are mainly caused by differing sea ice representations (Cubasch and Stössel 1994, personal communication). The high latitudes are weakly represented in V', however.

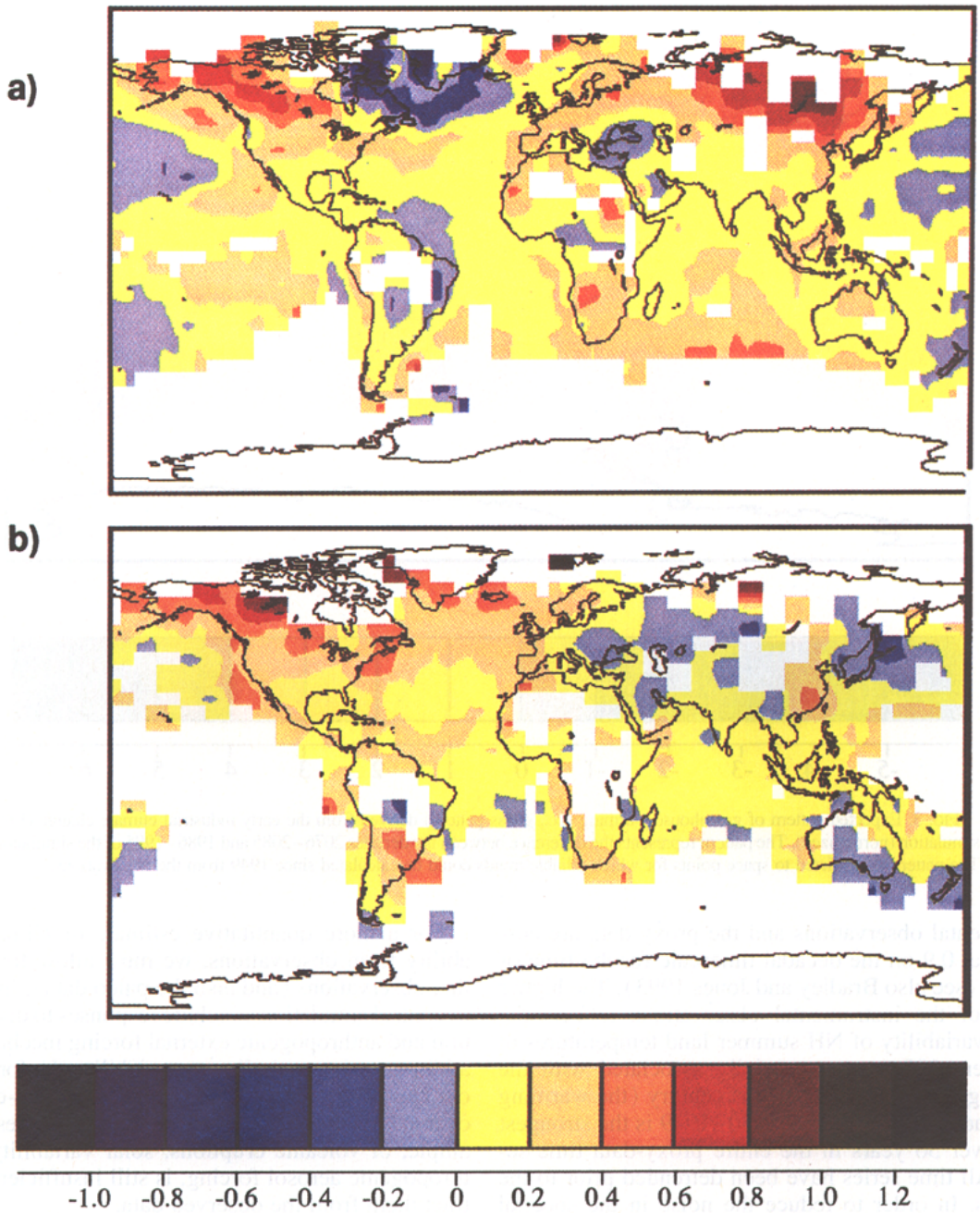

FIG. 2. Observed patterns of 30-yr trends for the periods 1965–1994 (a) and 1916–1945 (b) in degrees Celsius per decade, calculated from the data of Jones and Briffa (1992, 1994b).

Thus, our guess-pattern is a reasonable representation of the present model predictions of the spatial pattern of greenhouse warming.

d. Estimating natural variability

In order to estimate the statistics of the detection variable, numerous realizations $\tilde{\Psi}$ of natural variability are needed. We derive these realizations from the output of climate model control simulations and the time series of observations.

As a test of the order of magnitude of model variability, we compare a power-spectrum of a proxy time series for decadal mean North Hemispheric (NH) summer land temperatures from 1400 to 1970 (Bradley and Jones 1993) with the corresponding spectra of NH summer land means from the instrumental observations and from the model control simulations (Fig. 4). The

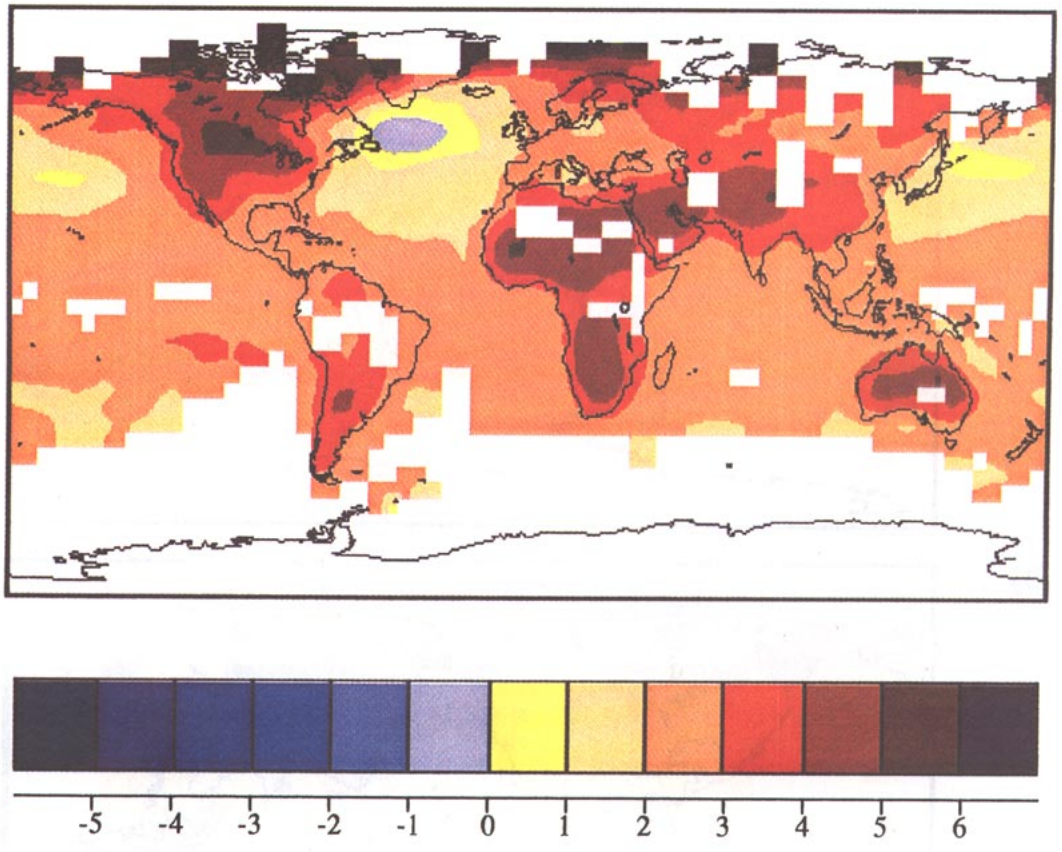

FIG. 3. Expected pattern of greenhouse warming (i.e., guess pattern) derived from the early industrial climate change (EIN) simulation (normalized). The pattern represents the difference between the decades 2076–2085 and 1986–1995 of the simulation. The pattern is restricted to space points for which reliable trends could be calculated since 1949 from the observations.

instrumental observations and the proxy-data are correlated at 0.9 on the decadal timescale for the time of overlap (see also Bradley and Jones 1993). The higher power of the instrumental observations reflects the higher variability of NH summer land temperatures in the recent 100 years, especially associated with the warming in the first half of this century (the warming trend ending in the decade 1930–1949 is the strongest trend over 30 years in the entire proxy-data time series). All time series have been detrended prior to the analysis. In order to reduce the noise in the spectral estimate, the time series are cut into segments of equal length (two segments of 65 years for the observations, five segments of 200 years for the long CTL simulations, two segments of 105 years for HAMO, and five segments of 11 decades for the first 550 years from the paleo data), the resulting spectra are then averaged. The results suggest that the variability of the climate models (for NH summer land mean temperature) on timescales of a few decades is quite different, but not inconsistent, with the observations and that the instrumental data provide a rather conservative (i.e., upper) estimate of natural climate variability on these timescales compared to the paleo data.

For a more quantitative estimate of climatic variability from observations, we must allow for the fact that observations (and also the paleo data), in contrast to control simulations, include responses to diverse natural and anthropogenic external forcing mechanisms in addition to internal climate variability. Unfortunately, our knowledge about the magnitude and time-evolution of many of these forcings and their responses, for example, of volcanic eruptions, solar variability, or anthropogenic aerosol forcing, is still insufficient to subtract them from the observed data.

However, we have a reasonably reliable record of CO_2 concentrations and can estimate the climate's response to the CO_2 increase from model simulations. We apply a linear response model to estimate and subtract the greenhouse warming pattern (Tahvonen et al. 1993) using only model-derived information about the greenhouse gas response. Similar response models have been used for global mean values by Sausen et al. (1988) and Barnett et al. (1991).

We note that the model estimate of the greenhouse warming signal, which is subtracted from the observations, depends on the estimates of the parameters of the model, for example, on the sensitivity to green-

TABLE 1. Winter and summer pattern correlations between climate change simulations with various CGCMs: the two climate models ECHAM/LSG (HAML), ECHAM2/OPYC (HAMO), and Hadley Centre model (UKMO), the NCAR model (NCAR), the GFDL model (GFDL) (for descriptions of the models see Gates et al. 1993), and the average over all model patterns. Climate change patterns are defined as the deviations from the control simulation of the decadal mean at the time of CO_2 doubling (except for NCAR that does not reach doubling). The upper diagonal gives results for the pattern correlation with spatial mean included [Eq. (8)], the lower diagonal for pattern correlation with the spatial mean subtracted.

	HAML	HAMO	UKMO	NCAR	GFDL	Mean
			Winter			
HAML		0.82	0.82	0.72	0.85	0.91
HAMO	0.56		0.86	0.80	0.92	0.95
UKMO	0.57	0.58		0.73	0.91	0.94
NCAR	0.37	0.46	0.35		0.82	0.86
GFDL	0.64	0.68	0.75	0.50		0.98
Mean	0.81	0.81	0.83	0.64	0.89	
			Summer			
HAML		0.64	0.66	0.67	0.72	0.82
HAMO	0.11		0.81	0.76	0.85	0.91
UKMO	0.27	0.42		0.73	0.89	0.92
NCAR	0.37	0.27	0.26		0.80	0.86
GFDL	0.32	0.44	0.67	0.36		0.96
Mean	0.59	0.65	0.80	0.55	0.82	

house gas forcing, and is thus subject to considerable uncertainty. If the models are erroneous, subtracting a wrong signal would introduce an additional artificial negative signal to the data, a procedure that should intuitively increase the variance. Also, the detection results do not change substantially if we simply use the observed trends without subtracting an estimate of the greenhouse gas signal. We prefer the present method to using only an early part of the observed data (which would presumably be less influenced by a developing greenhouse gas signal), since this would leave only a rather too short and spatially sparse record. Moreover, the choice of the "cut-off point" would necessarily be arbitrary.

To establish the linear climate response model, we assume, as in Cubasch et al. (1992), that the spatial pattern of the climate change signal is the dominant EOF $e_1(x, y)$ of the climate change simulation $s(x, y, t)$. Denoting the time-dependent evolution of the climate change signal by $r(t)$,

$$s(x, y, t) = r(t)e_1(x, y), \tag{5}$$

we assume that r evolves in time in accordance with the following simple time-dependent response model

$$\frac{\partial}{\partial t} r(t) = \mu C_f(t) - \alpha r(t), \tag{6}$$

where μ is a constant defining the sensitivity of the loading of the climate change signal to CO_2-forcing, and α is a memory term. The forcing of the model is given by the logarithmic relation (Callendar 1938)

$$C_f(t) = \ln\left[\frac{C(t)}{C_0}\right], \tag{7}$$

where $C(t)$ represents the CO_2 concentration that drives the model, and C_0 is the preindustrial concentration. The model is assumed to start from an equilibrium state (which is, of course, an idealization of the true climate evolution but is appropriate for the CGCM whose dynamics we wish to simulate).

The sensitivity and memory parameters are estimated by fitting the model (6) to the response $r(t)$ of the EIN simulation. Both the concentrations and $r(t)$ are filtered by an 11-yr running mean in order to remove high-frequency fluctuations. The resulting constants proved to be insensitive of the choice of the filter.

Figure 5 shows that the response $r(t)$ of the CGCM is well represented by the model with the parameters $\mu = 3.19$ C yr^{-1}, $\alpha = 0.0244$ yr^{-1}. Different constants, ($\mu = 8.54$ C yr^{-1}, $\alpha = 0.0359$ yr^{-1}), are obtained when the model is fitted to the observed time evolution of the climate change signal since 1860 in response to the observed CO_2 concentrations. The difference can be explained at least in part by the relatively slow onset of warming in the EIN-simulation (due to some combination of ice equilibration problems and internal variability; cf. Cubasch et al. 1995), possibly also by uncertainties in fitting the model to the observations due to gaps in the early data-sparse years.

We now drive the resulting response model with the observed CO_2 concentrations since 1860 (Keeling et al. 1989) and estimate the observed $r_{obs}(t)$ by integrating (6) with the forcing derived from the observed CO_2 concentrations since 1860. The computed temperature response field $r_{obs}(t)e_1(x, y)$ is then subtracted from the observed temperature patterns between 1860 and 1993. The residual field is our estimate of the natural

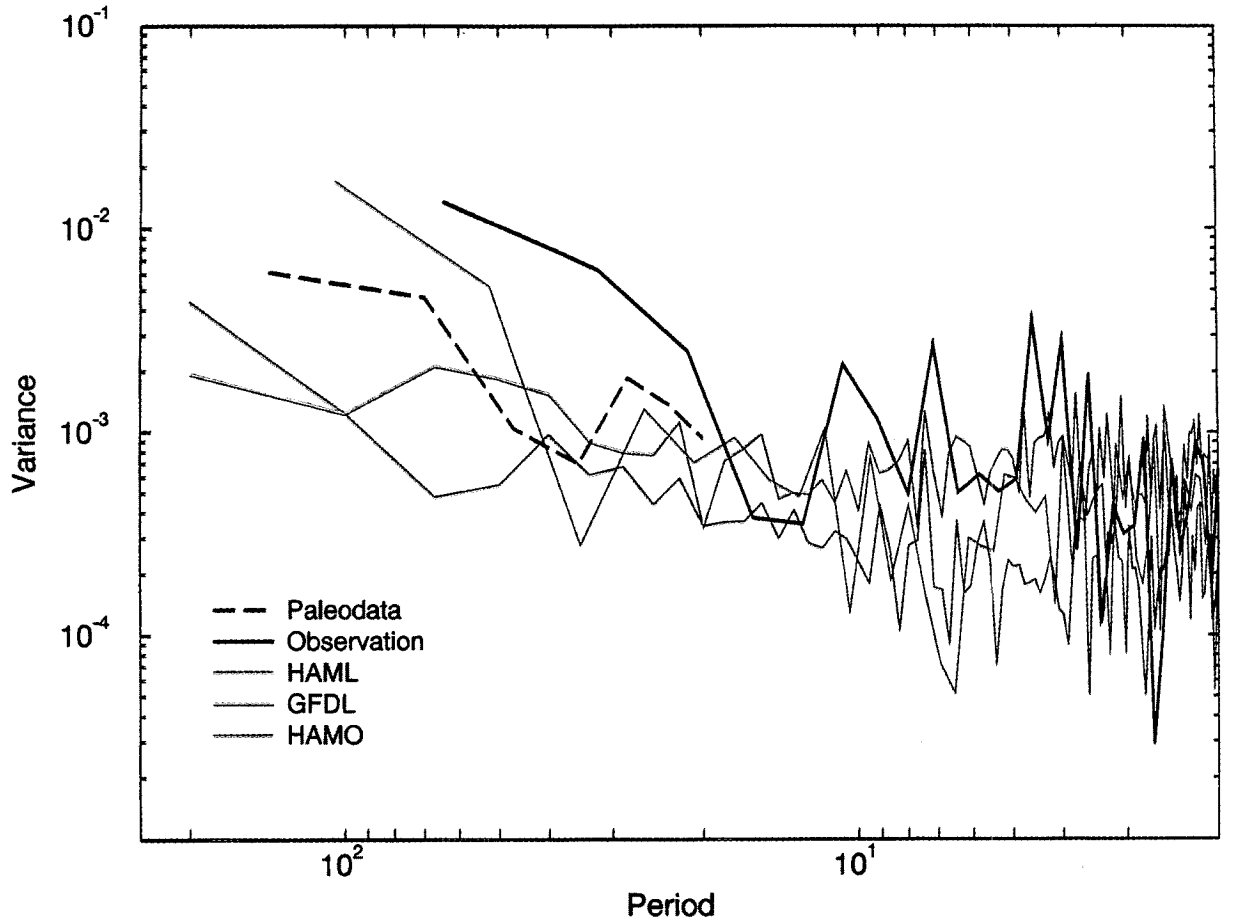

FIG. 4. Comparison of observed and simulated power spectra of mean Northern Hemisphere summer land temperatures for the instrumental observations, the model control simulations, and a proxy time series derived from paleoclimate data (Bradley and Jones 1993). Details in section 3d.

variability. It is possibly still contaminated by some residual greenhouse warming signal and by the response to external forcing other than CO_2, for example, aerosols or changes in solar insolation.

From the model control simulations and the observed variability data we have thus obtained four time series of patterns associated with natural variability data. The associated global means are shown in Fig. 1 (the dashed line in a is computed from the observations with the greenhouse gas subtracted). For each time t a spatial field of linear trends can be calculated between $t - l + 1$ and t, where l denotes the trend length (in analogy with the examples of trend patterns for the observed fields shown in Fig. 2). The projection of these data onto the detection space V yields time series of overlapping trend patterns, $\tilde{\Psi}_i(t)$, $i = 1, \cdots, 4$ of variability data. If the fingerprint is simply set equal to the guess pattern (no optimization), all four time series can be used for estimating the statistics of the detection variable.

To compute the optimal fingerprint, one of these natural variability time series needs to be selected to estimate the covariance matrix (3) of the climate noise. The other three variability time series can then be used to estimate the statistics of the detection variable, computed using the optimal fingerprint. The use of separate data for estimating the natural variability covariance matrix and the detection variable statistics is necessary to avoid a bias in the statistic, cf. section 2c.

We use the HAML control simulation for computing the variability covariance matrix. This choice has the following advantages:

- Figure 4 and the results presented below indicate that the internal variability of the HAML control simulation is smaller than both the observed variability and the variability of the HAMO and GFDL control simulations. Thus, an estimate of the statistics of the detection variable gained from the latter variability data will be larger than that obtained from the HAML data, yielding a more conservative detection test.

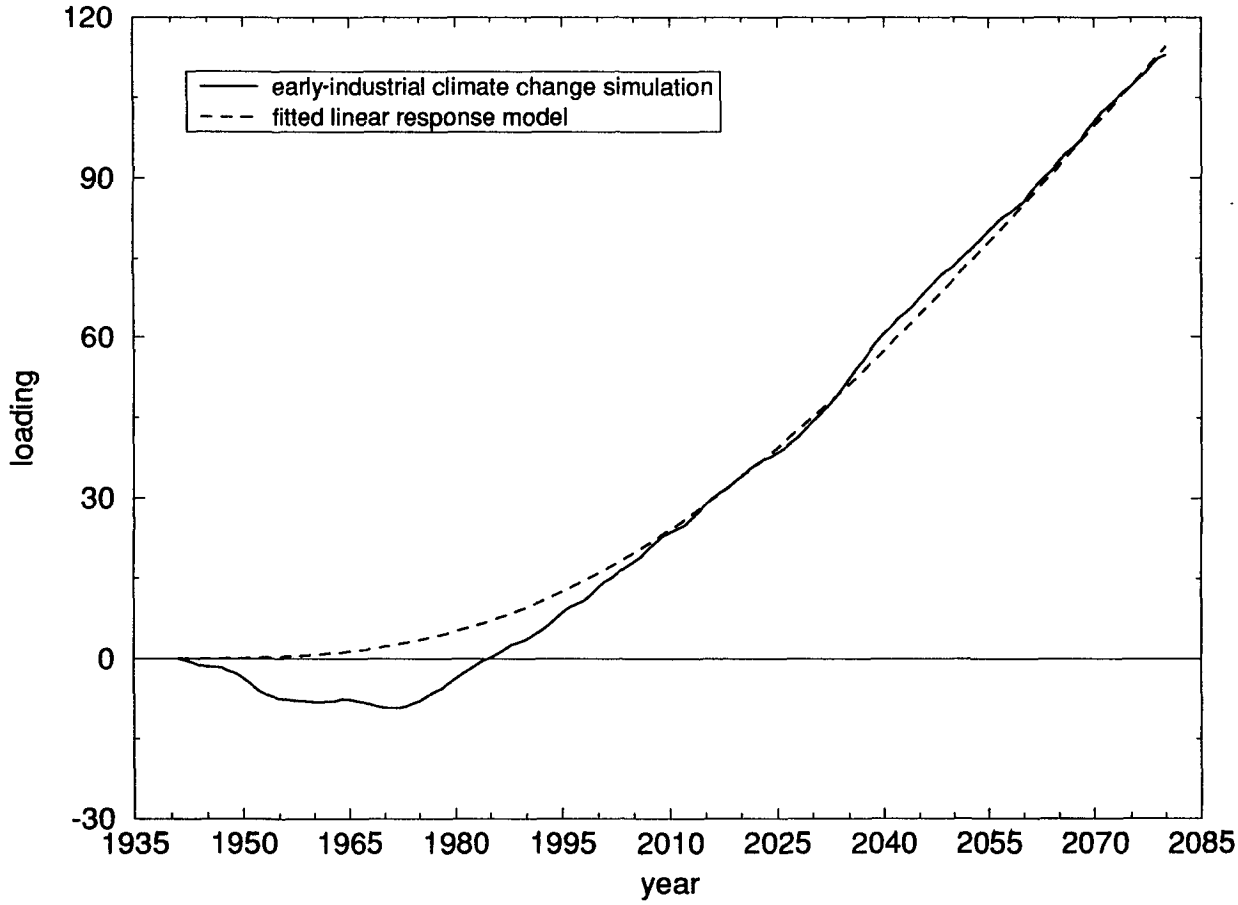

FIG. 5. Evolution of the climate change pattern (dominant EOF) for the EIN simulation (solid line). The dashed line represents the loading simulated by a linear response model fitted to the EIN response.

• The long (1000 yr) HAML control simulation yields enough samples for computing the optimal fingerprint (after truncation to a small dimensional space, see below).

e. Truncation to a small-dimensional space

Although we have a reasonably long record of variability data, the parameter space V' (consisting of about 1500 space points) needs to be further reduced in order to yield a reliable estimate of the noise covariance matrix. Note that the optimal fingerprint will only be optimal within the truncated subspace. The truncated space should be able to represent the signal of greenhouse warming sufficiently and represent at least the dominant features of climate noise.

Both criteria can be fulfilled by choosing the first few EOFs of a greenhouse warming simulation, for example, the EIN simulation. While the first EOF describes the time-evolving climate change signal, the following EOFs exhibit no clear trends and are associated with climate noise (see, e.g., Santer et al. 1994; Cubasch et al. 1995). The EOFs have been computed in the space V', disregarding the variance in regions not covered by observations.

The choice of the number of EOFs represents a compromise between a stable estimate of the noise covariance matrix (3) and the need for a sufficiently large space for optimizing the fingerprint. If we demand at least approximately three samples of 30-yr trends for each dimension of the covariance matrix, the dimension of the space should not exceed 10 since we have only 1000 years of noise data. Since we need to choose the truncation level a priori before computing the signal to noise ratios, we chose as a criterion the degree of rotation of the optimal fingerprint relative to the guess-pattern fingerprint. When the spatial dimension is too low, the rotation will be very small, indicating that there are not enough degrees of freedom for optimizing. If the truncation level is too high, the rotation will increase dramatically due to an underestimate of the variance associated with higher-indexed EOFs associated with undersampling (section 2c), leading to an artificial emphasis on small-scale noise. The correlation (in-

cluding the spatial mean) between the guess-pattern and the optimal fingerprint is a measure of the strength of rotation:

$$r = \frac{\mathbf{f}^T\mathbf{g}}{\|\mathbf{f}\|\,\|\mathbf{g}\|}. \quad (8)$$

Figure 6 shows the result for different truncation levels and 20-yr and 30-yr trends (results for 15-yr trends are not given since the rotation for these is small anyway; see section 4). After a moderate decrease of the pattern correlation with increasing number of EOFs (indicating stronger rotation due to more degrees of freedom), a more dramatic decrease occurs after eight EOFs for 20-yr trends and 10 EOFs for 30-yr trends. Figure 6 suggests that not more than eight EOFs should be used (we consider it safer to use the smaller estimate), later we see that the use of a different truncation level between four and 12 EOFs changes the results somewhat but does not yield different conclusions. However, we point out that errors in the determination of the optimal fingerprint direction arising from errors in the estimation of the variability covariance matrix (e.g. due to undersampling) result only in a conservative underestimation of the signal to noise ratio relative to the true (optimal) detection variable.

As truncated space V for the detection analysis we shall consider in the following the first eight EOFs of the EIN simulation. Subsequently, all data will be truncated in terms of this space. As expected, the truncation results in only a very small modification of the guess pattern: around 99.8% of the variance of the guess-pattern fingerprint in V' is recovered by V. To calculate the EOF-coordinates for observed data before 1949 (where grid points in V' may be missing), a least-square fit is used.

4. Results

a. *The spatial pattern of the optimal fingerprint*

The pattern correlation between the guess pattern and the optimal fingerprint in the reduced eight-dimensional detection space V is relatively high, especially for short trend lengths (0.98 for 15-yr trends and 0.82 for 30-yr trends, cf. Fig. 6). This is due to the fact that for the selected surface temperature data, the guess pattern is already fairly orthogonal to the HAML variability noise (see section 2a and Santer et al. 1994). Figure 7 shows the optimal fingerprint for 30-yr trends, transformed back from the EOF space V into the grid-point space V'. Compared to the original guess pattern (Fig. 3), the fingerprint pattern exhibits a stronger cooling in the upwelling areas of the North Atlantic and slight cooling even in the North Pacific. Cooling, or only slight warming, in this area due to changes in the ocean circulation is predicted in many greenhouse warming simulations (Cubasch et al. 1992, 1994, 1995). The optimal fingerprint emphasizes this pattern,

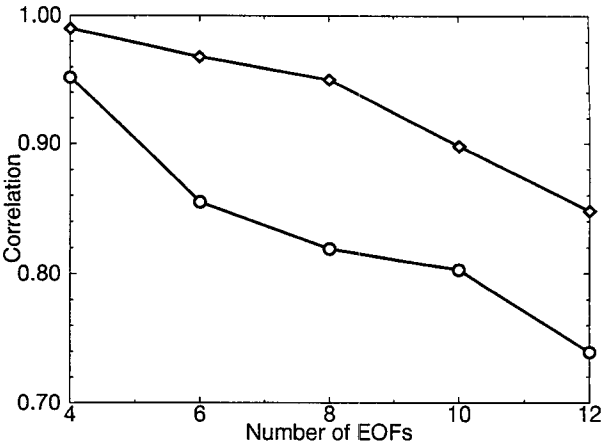

FIG. 6. Uncentered correlation [Eq. (8)] between the guess pattern and the optimal fingerprint (in the space spanned by the EOFs of the EIN simulation) as a function of the truncation level. Circles: 20-yr trends. Diamonds: 30-yr trends.

since it is associated with small natural variability (at least for our representations of natural variability). Additionally, the optimal fingerprint differs from the guess pattern over the large land masses associated with high natural variability.

The optimal fingerprint patterns for the 20-yr and 15-yr trends exhibit a similar, but smaller, enhancement of the cooling in the upwelling areas and less pronounced differences over the land masses. Optimal fingerprints computed from the other variability data also emphasized the cooling in the North Atlantic. However, for the optimal fingerprint computed from the covariance matrix estimate derived from the GFDL climate variability data, the cooling area was shifted westward to eastern North America. Nevertheless, the qualitative similarity of the optimal fingerprints calculated from the different natural variability data—but always using the same small reduced EOF phase space—gives some confidence that our estimate of the optimal fingerprint is reasonably stable.

b. *The evolution of the detection variable*

Figure 8 shows the time evolution of the detection variable $d(t) = \mathbf{f}^T\Psi(t)$ (where the time index t denotes the final year of each of the running trends) for the observations, computed with the fingerprint set equal to the guess pattern. Figure 9 shows the corresponding results for the optimal fingerprint based on the HAML variability data. Also plotted are the detection variables computed for the EIN simulation (dashed line). Before discussing the confidence intervals shown in the figures and testing the null-hypothesis that the latest observed trends (i.e., the 15-, 20-, and 30-yr trend ending in 1994) derive from our climate noise estimate, we point out some features of the detection variable time series.

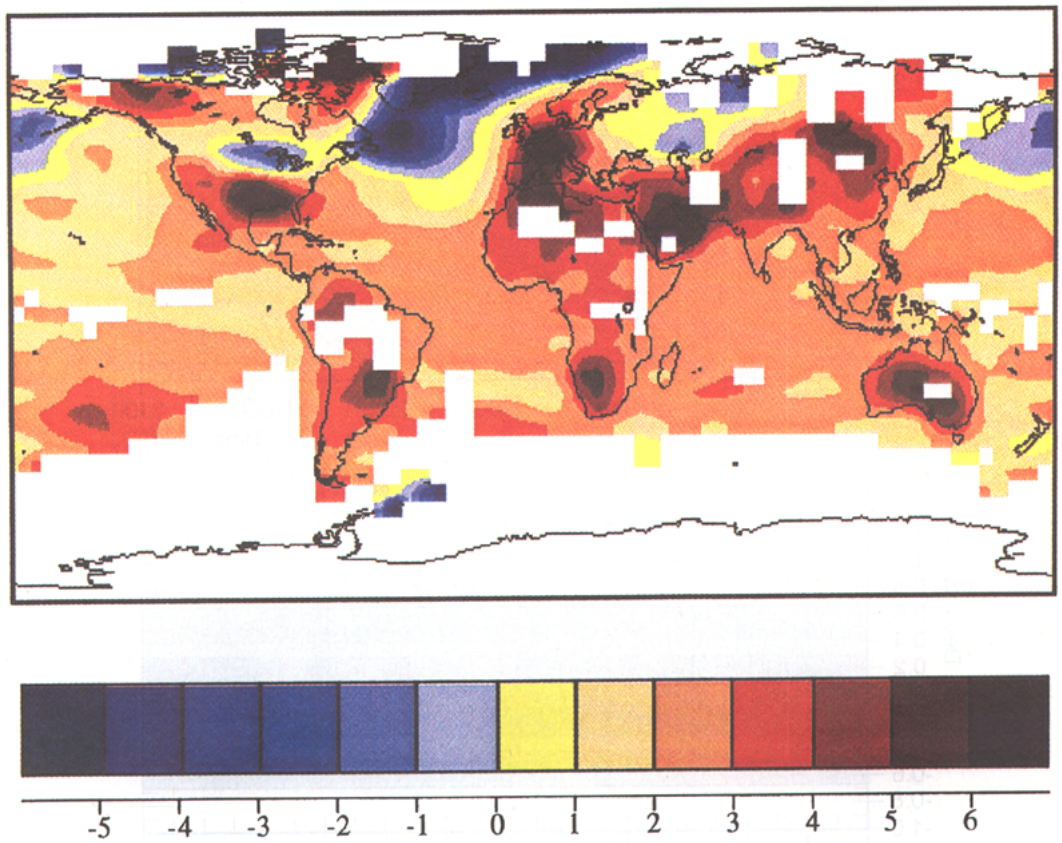

FIG. 7. Optimal fingerprint calculated from the HAML variability data for the 30-yr trends. The optimal fingerprint shows a more pronounced warming–cooling dipole in the North Atlantic than the guess pattern (Fig. 2).

• As expected, the evolution of the detection variable for short trend lengths is noisier than for longer (e.g., 30-yr) trends.

• The evolution of the detection variable in the observations is consistent with the model predictions. However, the warming in EIN is retarded relative to the observations. This may have several explanations. First, there is still some small cold start error in the EIN experiment, which does not have a true preindustrial start date (see Cubasch et al. 1995). Second, some of the slow initial rate of warming in the EIN integration may be due to ice equilibration problems: the spinup time of the coupled model was too short to yield a quasi-equilibrium state for the ice volume (see also Fig. 5). Third, the EIN experiment is only one realization of the response to the greenhouse gas forcing. Other realizations with the same greenhouse gas forcing, starting from slightly different (but equally plausible) initial conditions, yield different manifestations of the natural variability superimposed on the greenhouse warming signal. This leads to an uncertainty of approximately one decade in the onset of the accelerated warming (Cubasch et al. 1994). Further uncertainties are introduced by the restriction of the forcing to greenhouse gases only, and, of course, by systematic model errors.

The detection variable computed from the EIN simulation seems to increase until the middle of the next century and then fluctuate about a constant value. This indicates approximately constant temperature trends at this time, that is, a constant rate of temperature increase.

• For the 30-yr trends, the trend ending in 1945 (starting in the year 1916) is nearly as high as the latest observed trend (we stress, however, that we test the null hypothesis for the latest observed trend). During this period (1916–1945) a strong global mean warming lasting for about 30 years (cf. Fig. 1) was observed, which has been described, for example, by Parker et al. (1994). The NH summer land mean proxy-data (Bradley and Jones 1993) indicate that this was the strongest 30-yr warming trend observed over the last 500 years (see section 3d). Figure 2b shows that this warming differs from the present warming, especially in the high northern latitudes near Greenland, in Siberia, and in the interior of Australia. Part of this early warming may already be influenced by the climate change signal as estimated

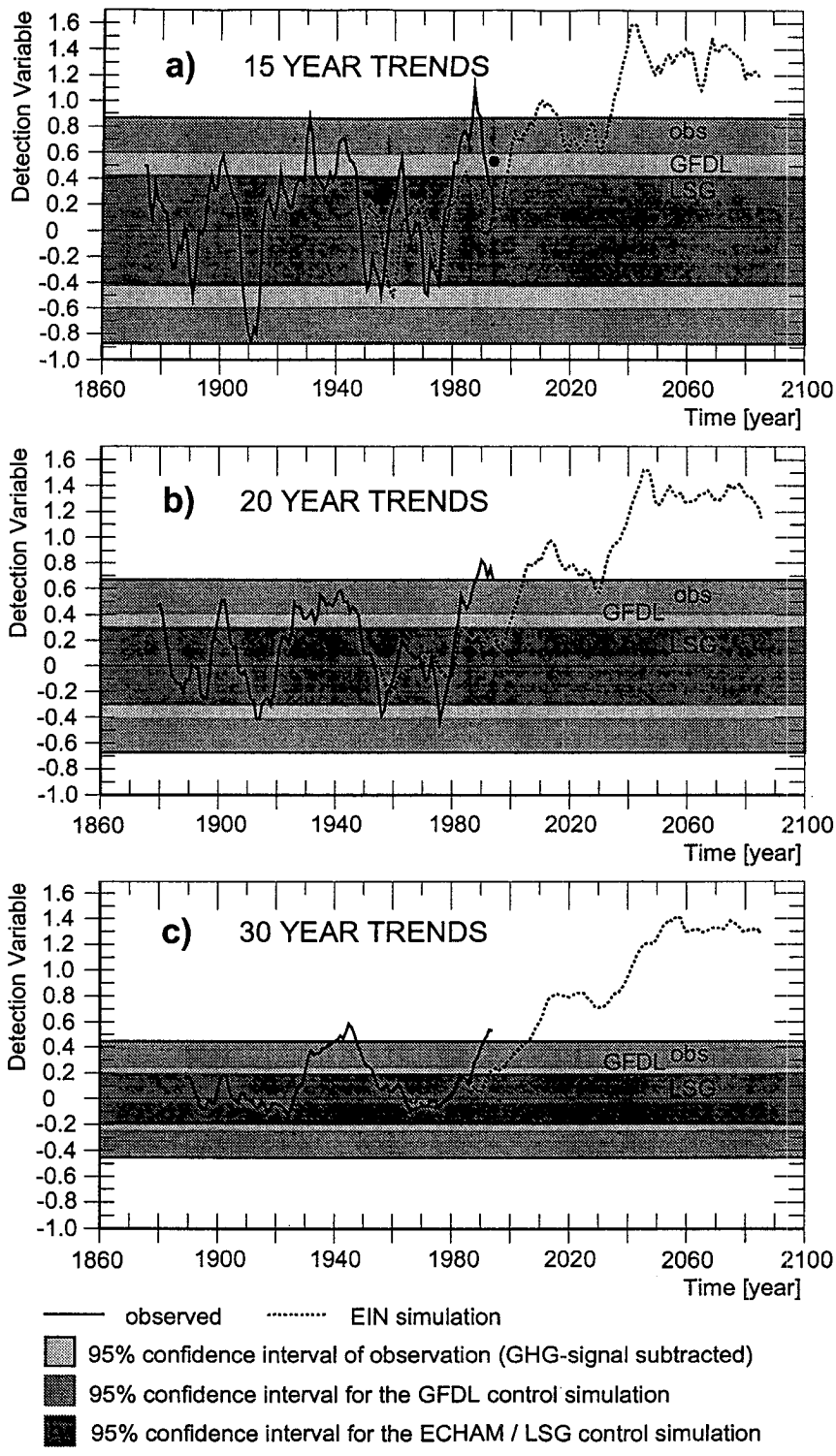

FIG. 8. Evolution of the detection variable using the guess-pattern fingerprint for the observations (solid line) and the EIN simulation (dashed line) for three trend lengths. The time refers to the final year of the trend. The dot at the year 1994 for the 15-yr trends represents the value after correcting for the cooling caused by the eruption of Mount Pinatubo and El Chichón (cf. section 4c). The 95% confidence intervals derived from three sets of variability data are indicated by different shading. For the present one-tailed test (the signal is known to be positive) the positive confidence limit corresponds to the 2.5% risk limit.

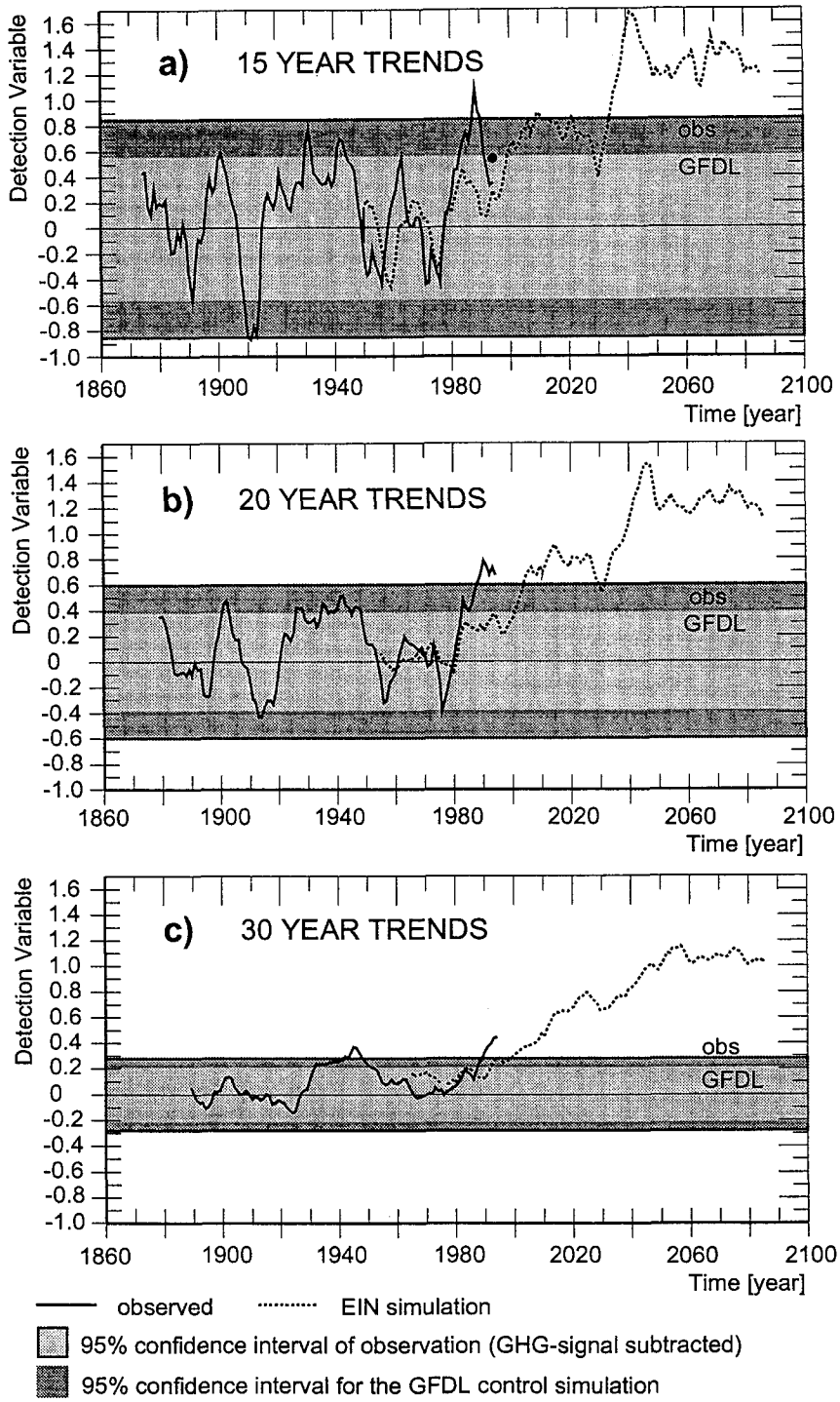

FIG. 9. Same as Fig. 8, but with the detection variable calculated using the optimal fingerprint. Since the optimal fingerprint has been computed from the HAML variability data, the 95% confidence intervals are shown only for the two other datasets.

from the EIN simulation (see Fig. 1a): the detection variable was reduced when the model-estimated greenhouse gas signal was subtracted from the observations (by 15% for the case that the fingerprint was set equal to the guess pattern and by 19% for the case of the optimal fingerprint).

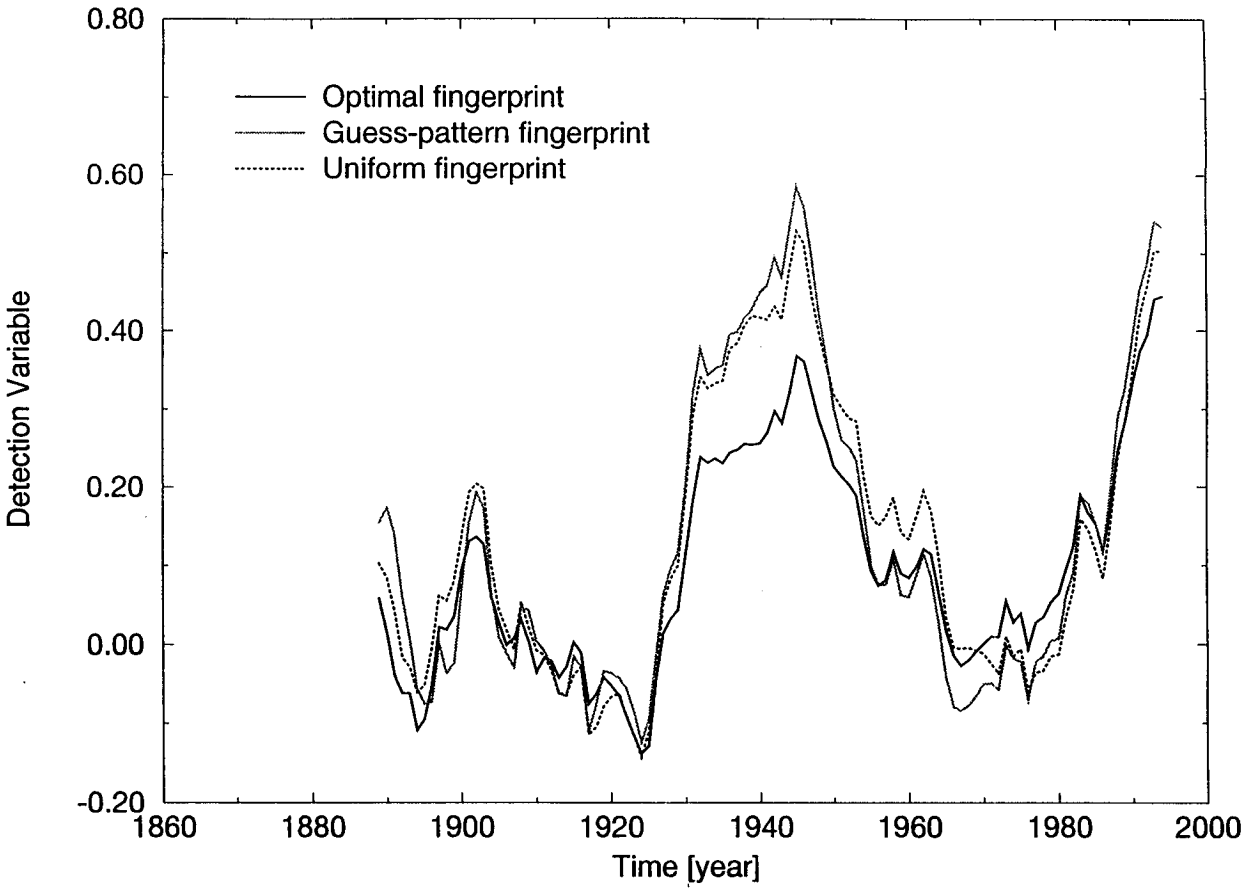

FIG. 10. Comparison between the detection variable calculated using the optimal fingerprint (full line), the guess-pattern fingerprint (gray line) and a spatially uniform fingerprint (dotted line) for 30-year trends.

• Since all fingerprint patterns are normalized to unity, the time evolution of the detection variable can be intercompared (section 2c). Figure 10 shows the evolution of the detection variable for the guess-pattern fingerprint compared to that of the optimal fingerprint for 30-yr trends. The absolute value of the detection variable is generally smaller for the optimal fingerprint than that for the fingerprint set equal to the guess pattern, but less so toward the end of the time series. This is due to suppression of climate noise by the optimal fingerprint, also evident in the wiggles of the time series of the detection variable. The strong warming trend ending in 1945 is decreased for the optimal fingerprint. Additionally, a detection variable based on a uniform pattern (reflecting global mean temperature trends) is shown. It is quite similar to the detection variable for the guess-pattern fingerprint. This illustrates our earlier comment that the main contribution to the detection variable comes from the global mean of the warming pattern (section 2c). The results of the optimal fingerprint, which is spatially less uniform (Fig. 7) disagree stronger from the global mean value. The results for 15- and 20-yr trends also show a close relationship between the guess-pattern fingerprint and the global mean change, the differences between the guess-pattern fingerprint and the optimal fingerprint results are more similar than for the 30-yr trends due to the greater pattern correlation.

c. *Significance of the latest observed trend*

The outcome of the statistical test depends critically on our estimate of the statistical properties of the detection variable for natural climate variability. This is based on four time series of overlapping trend patterns $\tilde{\mathbf{\Psi}}_i(t), i = 1, \cdots, 4$ for each trend length (section 3d). By computing the detection variable they are transformed into univariate time series $\tilde{d}_i(t)$.

We assume that the detection variable is Gaussian with zero mean (see section 2d), that is, that there is no long-term nonstationarity in the natural variability. The estimates of the standard deviations for each estimate of natural variability and for each trend length are listed in Table 2 for the guess-pattern fingerprint results and in Table 3 for the optimal fingerprint. The estimates differ markedly. While the differences might be partly

TABLE 2. Properties of the detection variable using the guess-pattern fingerprint for 15-, 20-, and 30-yr trends. The significance levels are calculated separately for each of the variability datasets: the observations with greenhouse gas signal subtracted (VOBS) and the control simulation of HAML, GFDL, and HAMO (cf. Table 1). The columns list (from left to right) the length of the time series $\tilde{d}(t)$, the decorrelation time T in years, the estimated standard deviation σ, the signal to noise variable S/N [Eq. (9)], and the risk for rejecting the null-hypothesis that the latest observed trend originates from natural variability based on a one-sided test [the risk levels are computed taking the number of samples and their serial correlation into account (cf. section 4c), thus they are higher than for a perfectly known Gaussian distribution]. Asterisks (*) denote that the null-hypothesis is not rejected at the 10% level.

Guess-f	Number of samples	T	σ	S/N	Risk
VOBS 15	120	22	0.40	0.85	*
HAML 15	986	21	0.21	1.65	10%
GFDL 15	986	18	0.30	1.13	*
HAMO 15	195	38	0.42	0.81	*
VOBS 20	115	27	0.30	2.23	2.5%
HAML 20	981	35	0.15	4.57	1%
GFDL 20	981	28	0.21	3.30	1%
HAMO 20	190	88	0.35	1.94	10%
VOBS 30	105	49	0.19	2.83	2.5%
HAML 30	971	61	0.095	5.61	1%
GFDL 30	971	48	0.13	4.25	1%
HAMO 30	180	219	0.28	1.94	10%

attributed to the limited sample size and the strong contribution to the variance from periods longer than the trend interval, each of the three estimates of the standard deviation of the detection variable for either fingerprint could also be biased for various reasons.

- The observed variability data (with the greenhouse gas signal subtracted) include the response to external forcing mechanisms unrelated to the greenhouse gas forcing (e.g., volcanism, changes in solar irradiance, and some anthropogenic aerosol signal) and, due to uncertainties in the subtracted anthropogenic climate response (section 3d), possibly also some residual greenhouse warming signal. These are not present in the control simulations and should intuitively lead to greater variability in the observations.
- Additionally, the first half of the observed variability data may be biased by poor spatial representation during the previous century and the early decades of this century. As mentioned in section 3b, if entire areas (and not just random gridpoints) are missing over longer periods, the detection variable from such sparse data might exhibit systematic errors. To estimate this effect, grid points for which there were no observations before 1900 were removed from the trend patterns of the HAML control simulation in V'. The standard deviation calculated from these reduced fields was about 10% larger than for the full spatial representation.
- The variability of both long control simulations with the HAML and the GFDL model is smaller than the observed variability. The HAML model yields about 50% of the standard deviation of the observations for all trend lengths, while the values for the GFDL model are between 75% (for 15-yr trends) and 67% (for 30-yr trends) of the observed level. To judge whether the differences could be attributed to external forcing in the observations, more must be known about the magnitude and spatial patterns of the climate response to such additional forcing mechanisms.
- The time series of the detection variable for the HAMO data is dominated by one single event representing a sudden cooling in the eleventh decade associated with a weakening of the thermohaline circulation (see section 3a). The simulation is too short to reliably estimate its variability excluding the extreme event. Also, contrary to the other variability estimates, the resulting standard deviation for the optimal fingerprint was found to be dependent on the truncation level (see, e.g., Table 4, section 4d). The results for the HAMO variability data are accordingly treated with caution, and the resulting confidence intervals are not shown in Figs. 8 and 9.

As expected, all standard deviation estimates for the optimal fingerprint (Table 3) are lower than for the guess-pattern fingerprint (Table 2) due to the suppression of noise by the optimization. The decrease for the 30-yr trends ranges between a factor of 0.61 (for HAMO) to 0.85 (GFDL); smaller decreases are obtained for the 15- and 20-yr trends.

We regard the differences of the estimated standard deviations for the different variability data as too large to compute a meaningful common estimate of σ from the pooled variability samples. We accordingly consider the statistical distributions separately.

As outlined in section 3d, the sample size is too small to use a Gaussian statistic with known variance for the statistic of the detection variable (especially for the ob-

TABLE 3. Same as Table 2 but for the optimal fingerprint computed from the HAML variability data. The columns list (from left to right) the pattern correlation r [Eq. (8)] between the guess pattern and the optimal fingerprint; T, σ, S/N same as in Table 2; and the ratio of S/N for the optimal fingerprint and the nonoptimized case (the results for the HAML variability data are not given since these data were used for calculating the optimal fingerprint).

Optimal f	r	T	σ	S/N	Risk	S/N Increase
15 VOBS	0.98	23	0.39	0.91	*	1.07
15 GFDL	0.98	19	0.28	1.24	*	1.10
15 HAMO	0.98	21	0.40	0.88	*	1.09
20 VOBS	0.95	27	0.27	2.46	2.5%	1.10
20 GFDL	0.95	27	0.19	3.46	1%	1.05
20 HAMO	0.95	75	0.30	2.27	5%	1.17
30 VOBS	0.82	36	0.12	3.66	1%	1.29
30 GFDL	0.82	49	0.11	4.01	1%	0.94
30 HAMO	0.82	118	0.17	2.66	2.5%	1.37

TABLE 4. Sensitivity of the signal to noise levels S/N with respect to the choice of truncation level (in terms of the number of EOFs of the EIN simulation used) and data used for the optimization. Results are listed only for 20- and 30-yr trends (the 15-yr trends are only marginally significant and the rotation in that case is weak). The right column gives the results if the data from the GFDL CTL simulation for are used for estimating the covariance matrix. An asterisk (*) denotes a decrease in signal to noise ratio due to the optimization.

Optimal f	S/N:4	S/N:6	S/N:8	S/N:10	S/N:12	GFDL, 10
20 VOBS	2.40	2.49	2.46	2.33	2.33	2.28
20 HAML						4.23*
20 GFDL	3.48	3.48	3.46	3.17*	3.17*	
20 HAMO	2.16	2.25	2.27	1.96	2.01	1.94*
30 VOBS	3.36	3.76	3.66	3.62	3.24	2.21*
30 HAML						4.04*
30 GFDL	4.44	4.17*	4.01*	3.74*	3.76*	
30 HAMO	2.44	2.61	2.66	2.27	2.28	1.90*

served variability). We use Monte Carlo simulations for estimating the error associated with estimating the variance from a finite-length, autocorrelated time series. Due to the additional uncertainty introduced by the variance estimate, the confidence interval will exceed that for a Gaussian distribution of the same but known standard deviation. We assumed that the $\tilde{d}(t)$ time series can be approximated by an AR(1) process. Several 10 000 experiments were made in order to estimate the statistical distribution of the signal-to-noise variable

$$S/N = d/\sigma, \tag{9}$$

in which the standard deviation was estimated from a random AR(1) time series whose length and autocorrelation scale was computed from the $\tilde{d}(t)$ time series.

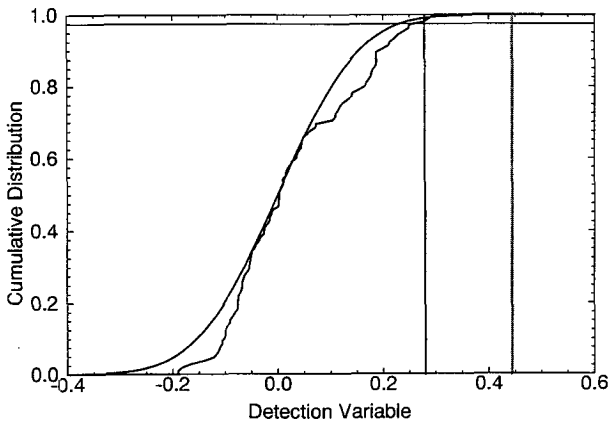

FIG. 11. Cumulative sampling distribution of the detection variable for the optimal fingerprint for 30-yr running trends from the observations with greenhouse gas signal subtracted. The smooth line indicates the Gaussian distribution with the same standard deviation. In this case the sampling distribution is derived from very few independent samples. The value of the detection variables at the intersection of the horizontal lines with the sampled and Gaussian distributions defines the 97.5% significance level, while the vertical line represents the 97.5% confidence level estimated from Monte Carlo simulations. The value of the detection variable for the latest observed temperature trend pattern is indicated by the gray vertical line.

The 95% two-sided confidence intervals based on the Monte Carlo simulations are shown in Figs. 8 and 9, they are higher than 1.96σ (perfectly known Gaussian noise, section 2d), for example, 2.4σ for 30-yr trends and σ estimated from the observed variability.

Figure 11 shows the cumulative distribution function for an ideal Gaussian distribution compared to the empirical distribution functions calculated from the 30-yr samples derived from the observed variability. The sampling distribution in this case is derived from extremely few independent samples, thus, the disagreement between the two distributions is not surprising. The skewness to positive values of the detection variable might indicate a residual greenhouse warming signal in the observed data (see also Fig. 1). The intersection of the curves with the value of 0.975 at the vertical axis defines the (one sided) 97.5% confidence limit of the distribution. The vertical line determines the 97.5% confidence level estimated from the Monte Carlo tests. In the cases shown and also in all other cases, the Gaussian and Monte Carlo simulated confidence limits were conservative (i.e., higher) relative to the use of the empirical sampling distribution. For the 1000-yr simulations, the estimated distribution functions were nearly indistinguishable from those for the ideal Gaussian distribution. We would like to emphasize that there is substantial uncertainty in the estimate of the statistics. However, Fig. 11 and Table 3 indicate that the detection variable for the optimal fingerprint is well beyond the 2.5% significance level for the latest observed trend, indicating that the uncertainty seems not to be critical in that case.

For all trend lengths, the detection variable for the latest few observed trends extends farther from the noise (as expressed by the confidence intervals in Figs. 8 and 9) in the case of optimal fingerprint than for the guess-pattern fingerprint case. The benefit of rotating the fingerprint can be expressed by the ratio of the signal to noise ratio S/N (9) for the optimal fingerprint to that for the guess-pattern fingerprint (cf. right column in Table 3). The increase is highest for the 30-yr trends ($\sim$1.3), which exhibit the strongest rotation expressed

by the smallest pattern correlation. An exception are the GFDL variability data (0.94). Although the variance of the natural variability data is also decreased in this case through application of the optimal fingerprint, this effect is too small to overcome the decrease of the amplitude of the detection variable. This results in a net decrease of S/N for the GFDL model (cf. section 2c) that may be associated with a different variability structure of the GFDL model compared with the HAML model (which was used for optimizing and for truncation). A different reaction of different variability data on optimization was also found in a model study (Santer et al. 1995b).

For all variability estimates, increases in S/N ranges from 1.1 to 1.2 for the 15-yr and 20-yr trends. In summary, optimization generally leads to an improvement in S/N rates.

The results of a one-sided statistical test of the null-hypothesis that the latest observed trend is associated with natural variability, as expressed in each dataset for natural variability, are summarized in the Tables 2 and 3 (column "risk"). The latest 15-yr trend is relatively insignificant, in contrast to the 15-yr trend ending in 1988 (Figs. 8, 9). We suspect that the lower atmospheric cooling in 1992 and 1993 caused by the eruption of Mount Pinatubo in 1991 has reduced the latest 15-yr trends, whereas it had less influence on the longer trends. Since the detection variable contains a strong component of the spatial mean (see section 5), the effect of Pinatubo and El Chichón on the latest 15-yr trends can be approximately estimated by subtracting the estimated global mean of the volcano-induced temperature change (see Jones 1994a) from each grid point. The resulting value of the detection variable for the latest observed trend is indicated by the dots in Figs. 8 and 9.

The detection variable (for both the guess-pattern fingerprint and the optimal fingerprint) for the latest 20-yr trend deviates from natural variability as estimated from the GFDL and observed variability data with an estimated risk of less then 2.5%. In the case of the HAMO variability, the risk is 5% for the optimal fingerprint, as compared with 10% for the guess-pattern fingerprint.

For the 30-yr trends the detection variable deviates from natural variability as estimated from the GFDL and observed variability data with an estimated risk of less then 2.5% for the guess-pattern fingerprint and only 1% for the optimal fingerprint. For the HAMO variability the risk is less than 2.5% for the optimal fingerprint as compared with 5% for the guess-pattern fingerprint.

d. Discussion

Our estimate of the significance levels are dependent on a number of assumptions and conditions.

- It is assumed that our estimate of the statistics of the detection variable is correct, that the detection variable is Gaussian and that in estimating its variance, the time series can be modeled approximately by an AR (1) process. We consider the uncertainty introduced by this assumption small compared to the limitations of the natural variability data.
- The results are not sensitive to the choice of variability datasets or truncation level used for optimizing: the dependencies are indicated in Table 4. The central column (bold face) corresponds to the standard case (eight EOFs, optimization with the HAML control simulation data, cf. section 3e). The left–right columns in Table 4 give the results if the optimization is performed in a smaller higher dimensional space (spanned as before by EOFs of the HAML model EIN simulation). As expected from Fig. 6, the results deteriorate for 10 or 12 EOFs, but the results of the statistical test (i.e., the significance levels) change only in the case of the HAMO simulation. The optimal fingerprint was also computed from the GFDL data (using a truncation to 10 EOFs of the EIN simulation). The results in this case are not as well-behaved, the signal to noise variable decreasing relative to the nonoptimized case for all 30-yr trend data and for some data for the 20-yr trends. This indicates differences in the variability structure between the GFDL and the HAML model.
- It is assumed that the variability data used for estimating the statistics of the detection variable represent the natural variability of climate correctly. This is the biggest uncertainty of our study, due to the uncertainty as to whether the models simulate climate variability realistically on decadal timescales, the relatively short and spatially sparse time series of the observations, and the procedure of subtracting the greenhouse warming signal from the observations. However, even without the subtraction of the greenhouse warming signal from the observations, the latest 20-yr and 30-yr trends are found to be significant at the 5% level relative to the observed variability in such trends since 1860. More work with paleoclimatic data may be useful in assessing the reliability of model noise on the decadal to century timescales.

5. Attribution

Although we have been able to statistically identify a climate change signal at an estimated 2.5% significance level (with uncertainties inherent in the estimate of natural climate variability), we have not yet established that this signal is in fact caused by an increase in greenhouse gas concentrations.

To attribute the observed unusual warming to anthropogenic greenhouse gas forcing, we must rule out the possibility that other natural or anthropogenic forcing mechanisms, such as volcanic activity, changes in solar radiation, aerosols, or a superposition, together with possible nonlinear interactions of these mecha-

nisms, could cause a similar response of the climate (Santer et al. 1993a). We do not consider these alternative forcing mechanisms in detail, although most appear improbable.

- Decadal changes in solar radiation are generally believed to cause only a small response of the climate (Hansen and Lacis 1990). However, longer timescale solar variability may be stronger. Estimates of solar variability indicate an increase of solar radiation in the first half of this century and since about 1960 (Hoyt and Schatten 1993). More reliable information on changes in solar radiation and the climate's response pattern to it is needed to provide an estimate of the influence of solar variability on the observed warming.
- Major volcanic eruptions are relatively infrequent and cause a decrease of the global mean near-surface temperature on the timescale of only a few years. Although it has been speculated that the clustering of volcanic eruptions may influence the temperature variability on longer timescales, it is difficult to see how recent volcanic eruptions could have induced the observed global warming signal. For an overview over both solar and volcanic forcing influences see Crowley and Kim (1993).
- The direct albedo effect of anthropogenic aerosols alone causes a net global cooling and can therefore obviously be ruled out as a global warming candidate. However, recent climate change simulations including both anthropogenic greenhouse gas and aerosol forcing yield modified signal patterns. In two rather similar transient simulations with CGCMs, Mitchell et al. (1995) and Hasselmann et al. (1995) found that the warming in the north hemispheric midlatitudes is reduced due to the aerosol cooling. The same features were found in analogous equilibrium AGCM experiments; Santer et al. (1995a) found an enhanced correlation between the simulated and observed pattern of surface temperature changes in summer and autumn, the pattern correlation increasing with time. They conclude that at present an enhanced greenhouse warming signal alone may be difficult to detect in observed temperature data by a centered pattern correlation statistic since anthropogenic sulfate aerosols may have partially obscured the greenhouse warming signal.

We would have more confidence that the observed unusual warming can be uniquely attributed to anthropogenic greenhouse gas forcing if the observed warming would agree with the greenhouse gas signal also in terms of pronounced features of its the smaller-scale spatial structure. This is, however, not the case. The single most pronounced feature of the observed climate change is the rise in global mean temperature. Figure 12 shows the (centered) pattern correlation between the guess-pattern fingerprint and the patterns of observed 20- and 30-yr trends. Although the pattern correlation between the guess-pattern fingerprint and the observed trend patterns is positive (around 0.3 for recent 20- and 30-yr trends), it is relatively small. This agrees qualitatively with previous studies by Santer et al. (1993b, 1995a) in which the correlation between various model-predicted equilibrium signal patterns (in response to present CO_2 concentrations) and observed time-dependent patterns of near-surface temperature change showed no evidence of sustained positive trends over recent decades. There was thus no evidence that the mean-subtracted signal patterns (predicted by five different earlier GCMs without coupling to ocean GCMS) were becoming more evident in the observed data as the global mean temperature increased.

Figure 12 also shows that the correlations between the observed and model-simulated trend patterns in our case are nevertheless consistent with the EIN model prediction. Contrary to the 20-yr trends, which appear too noisy to sustain a high spatial correlation, the model predicts that the 30-yr trends will increase further if the warming has in fact been caused by the greenhouse gas forcing alone. However, we have not attempted here to correct for the contamination by aerosol forcing. In this sense our fingerprint is presumably still rather far from optimal.

In summary, although the observed climate change is consistent with the model prediction and we offer no other convincing explanation for the climate change, we cannot at this stage uniquely attribute the observed climate change to the greenhouse gas forcing. To rigorously rule out other possible climate change mechanisms, these need to be specified and compared with greenhouse forcing, for example, using a multiple fingerprint technique (Hasselmann 1993). Preliminary results of a two-fingerprint approach using a CO_2 only and a CO_2 + Aerosol optimal fingerprint show that such a method is able to distinguish between two rather similar patterns (in terms of the dominant annual mean climate change signal; Hasselmann et al. 1995) and that the observations agree substantially better with the combined forcing pattern.

6. Conclusions

We have shown that the use of a statistically optimal fingerprint suppresses natural variability noise as represented by both model internal variability and observed climate variability and generally enhances the signal to noise ratio. We find that the latest observed 30-yr trend pattern of near-surface temperature change can be distinguished from all estimates of natural climate variability with an estimated risk of less than 2.5% if the optimal fingerprint is applied. This also holds for the latest 20-yr trend with the exception of the HAMO variability data (which has problems, however, see below). The latest observed 15-yr trend, in contrast to the maximal 15-yr temperature trend between 1974 and 1988, is not unusually strong. This is in part due to the global cooling caused by the eruption of Mount Pinatubo and to noise contamination, both of which have a stronger influence on shorter timescales.

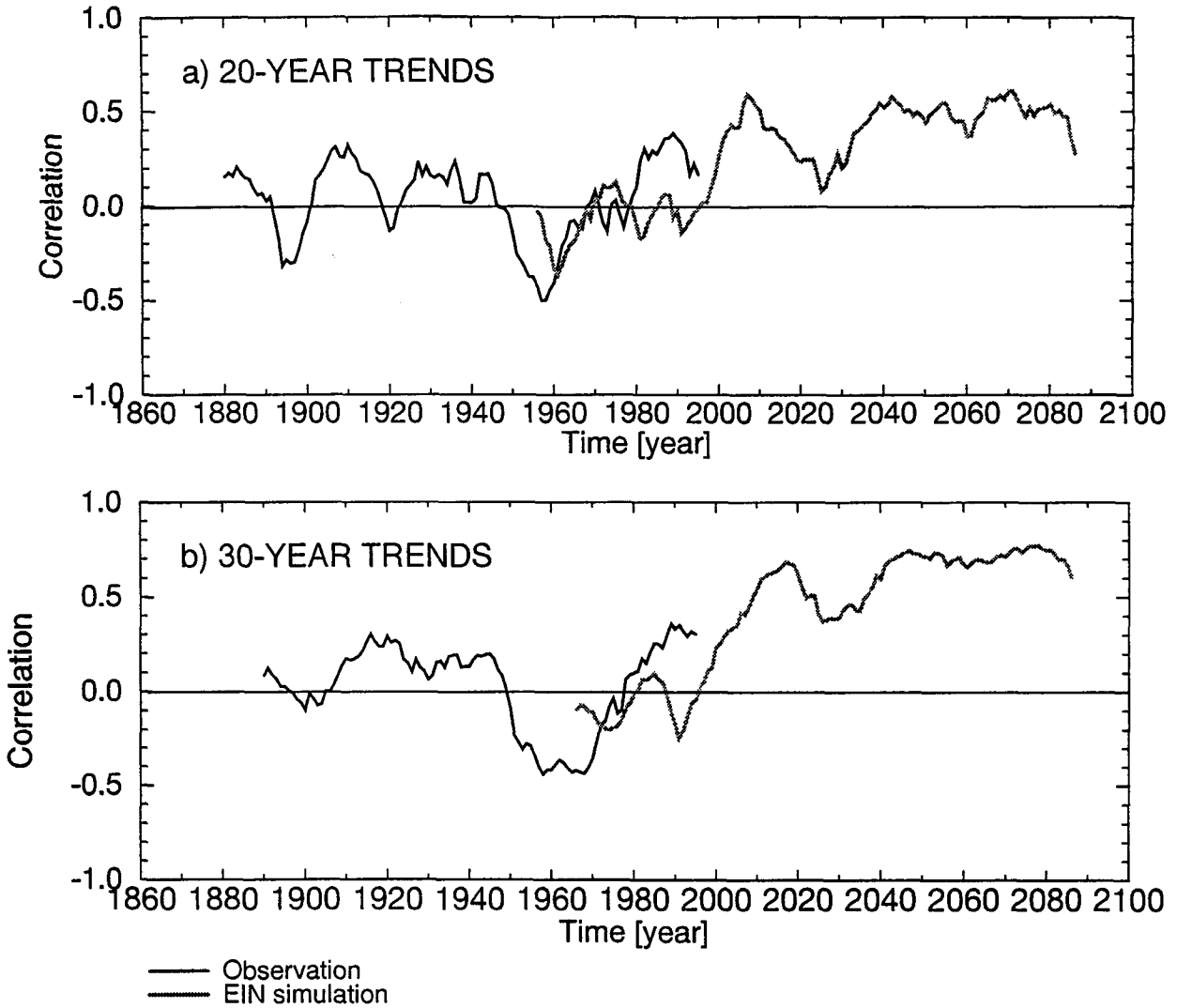

FIG. 12. Centered pattern correlation between the guess pattern and the observed 20- and 30-yr trend patterns (full line) and the trend patterns computed from the EIN simulation (gray line).

The greatest uncertainty of our analysis is the estimate of the natural variability noise level. Both the observed and model simulated datasets used to estimate the internal climate variability have limitations.

- The observational data are short (~140 years) and very sparse in the early years. Furthermore, even after subtraction of the estimated greenhouse warming contribution, the observations may still incorporate the effect of a residual greenhouse warming signal and other external forcing mechanisms (e.g., volcanic and anthropogenic aerosols, solar variability).
- The variability in decadal to century timescales of the HAML simulation seems to be smaller than that of the GFDL simulation; both models exhibit less variability than the observations.
- The variability in the HAMO simulation is strongly affected by a single, possibly unrealistically strong event associated with a change in the ocean circulation. Because of the relatively short integration time (210 years), this event dominates the estimate of natural variability; we treat the variability estimate based on this simulation with caution.

Furthermore, the limited length and high autocorrelation of the time series of the detection variable derived from the natural variability data introduces sampling uncertainties in the estimate of the natural variability of the detection variables. To correct for sampling errors in estimating the natural variability of the detection variable from short time series, such as the observations, we use Monte Carlo simulations. The

underlying assumptions of this approach are that the noise is Gaussian and generated by an AR (1) process. Although this is undoubtedly a simplification, we regard the sampling uncertainties as small compared with the other uncertainties of the variability data.

There are some indications that the variability in the instrumental record is greater than the variability earlier in the paleo-time series (Fig. 4). This relationship suggests that a variability estimate based on the instrumental period yields a conservative estimate of the signal to noise ratio. This gives us some added confidence that a significant warming has in fact been observed.

The shortcomings of the present estimates of natural climate variability cannot be readily overcome. However, the next generation of models should provide us with better simulations of natural variability. In the future, more observations and paleoclimatic information should yield more insight into natural variability, especially on longer timescales. This would enhance the credibility of the statistical test. Moreover, the results of the EIN simulation, as well as new simulations with CO_2 and aerosol forcing (Mitchell et al. 1995; Hasselmann et al. 1995), indicate that the signal of anthropogenic climate change is now expected to emerge from the background noise. Thus, although we shall never be able to overcome the basic uncertainty that any detection approach depends upon an *estimate* of natural variability whose reliability cannot be assessed rigorously, the next years might add to our confidence that an anthropogenic climate change signal is contained in the observations.

Although we consider anthropogenic greenhouse gas forcing to be the most likely candidate, we cannot yet uniquely attribute the observed abnormal warming to this mechanism. The main contribution to our detection variable comes from the increase of the global mean temperature. Model simulations of the climate response patterns for various competing external forcing mechanisms (and their possible nonlinear interactions) combined with multiple-fingerprint detection methods are needed to enhance our confidence that the present warming has indeed been caused by anthropogenic greenhouse gas forcing.

Acknowledgments. We thank Tom Crowley, Mark Berliner, Arno Hellbach, Heinke Höck, Frank Lunkeit, Ernst Maier-Reimer, Uwe Mikolajewicz, Josef M. Oberhuber, Hinrich Reichardt, Erich Roeckner, Victor Ocana, Ralf Weisse, Dierk Schriever, and Eduardo Zorita for their support in the scientific and practical aspects of this work and Ray Bradley for the paleoclimatic data. M. Grunert drew the diagrams. We are also grateful to G. North and an anonymous reviewer who helped us to improve the paper. The research has been supported by the German Ministry for Research and Technology (BMFT) and the EC "Environment" program under the Contract No. EVSV-CT92-0123. The authors are grateful to the staff of the DKRZ for their technical support.

REFERENCES

Barnett, T. P., 1986: Detection of changes in global tropospheric temperature field induced by greenhouse gases. *J. Geophys. Res.*, **91,** 6659–6667.

——, 1991: An attempt to detect the greenhouse-gas signal in a transient GCM simulation. *Greenhouse-Gas-Induced-Climatic Change: A Critical Appraisal of Simulations and Observations,* M. E. Schlesinger, Ed., Elsevier Science, 559–568.

——, and M. E. Schlesinger, 1987: Detecting changes in global climate induced by greenhouse gases. *J. Geophys. Res.*, **92,** 14 772–14 780.

——, 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. Schlesinger, 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. *The Holocene,* in press.

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

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

Bradley, R. S., and P. D. Jones, 1993: 'Little Ice Age' summer temperature variations: Their nature and relevance to recent global warming trends. *The Holocene,* **3,** 367–376.

Briffa, K. R., and P. D. Jones, 1993: Global surface air temperature variations during the twentieth century: Part 2, implications for large-scale high-frequency palaeoclimatic studies. *The Holocene,* **3,** 77–88.

Callendar, G. S., 1938: The artificial production of carbon dioxide and its influence on temperature. *Quart. J. Roy. Meteor. Soc.,* **64,** 223–235.

Crowley, T. J., and K-Y. Kim, 1993: Towards development of a strategy for determining the origin of decadal-centennial scale climate variability. *Q. Sci. Rev.*, **12,** 375–385.

Cubasch, U., K. Hasselmann, H. Höck, E. Maier-Reimer, U. Mikolajewicz, B. D. Santer, and R. Sausen, 1992: Time-dependent greenhouse warming computations with a coupled ocean-atmosphere model. *Climate Dyn.*, **8,** 55–69.

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

——, G. C. Hegerl, A. Hellbach, H. Höck, U. Mikolajewicz, B. D. Santer, and R. Voss, 1995: A climate change simulation starting from 1935. *Climate Dyn.*, **11,** 71–84.

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

Dickson, R. R., J. Meincke, S. A. Malberg, and A. J. Lee, 1988: The "great salinity anomaly" in the northern North Atlantic 1968–1982. *Progress in Oceanography,* Vol. 20, Pergamon, **20,** 103–151.

Folland, C. K., T. R. Karl, N. Nicholls, B. S. Nyenzi, D. E. Parker, and K. Y. Vinnikov, 1992: Observed climate variability and change. *Climate Change 1992. The Supplementary Report to the IPCC Scientific Assessment.* J. T. Houghton, B. A. Callander, and S. K. Varney, Eds., Cambridge University Press, 135–170.

Gates, W. L., J. F. B. Mitchell, G. J. Boer, U. Cubasch, and V. P. Meleshko, 1992: Climate modelling, climate prediction and model validation. *Climate Change 1992. The Supplementary Report to the IPCC Scientific Assessment.* J. T. Houghton, B. A. Callander, and S. K. Varney, Eds., Cambridge University Press, 97–134.

——, U. Cubasch, G. A. Meehl, J. F. B. Mitchell, and R. J. Stouffer, 1993: An intercomparison of selected features of the control climates simulated by coupled ocean–atmosphere general cir-

culation models. Tech. Rep. WMO/TD 574, 46 pp. [Available from WMO, Case Postale 2300, CH-1211 Geneva 2.]

Hannoschöck, G., and C. Frankignoul, 1985: Multivariate statistical analysis of sea surface temperature anomaly experiments with the GISS general circulation model. *J. Atmos. Sci.,* **42,** 1430–1450.

Hansen, J. E., and A. A. Lacis, 1990: Sun and dust versus greenhouse gases: An assessment of their relative roles in global climate change. *Nature,* **346,** 713–719.

Hasselmann, K., 1979: On the signal-to-noise problem in atmospheric response studies. *Meteorology over the Tropical Oceans,* D. B. Shaw, Ed., Roy. Meteor. Soc., 251–259.

——, 1993: Optimal fingerprints for the detection of time-dependent climate change. *J. Climate,* **6,** 1957–1971.

——, and Coauthors, 1995: Detection of anthropogenic climate change using a fingerprint method. MPI Rep. 168, 20 pp. [Available from Max-Planck-Institut für Meteorologie, Bundesstr. 55, 20146 Hamburg, Germany.]

Houghton, J. T., and L. G. Meira Filho, 1996: *Climate Change 1995. The IPCC Second Scientific Assessment.* Cambridge University Press, 572 pp.

——, G. J. Jenkins, and J. J. Ephraums, 1990: *Climate Change. The IPCC Scientific Assessment.* Cambridge University Press, 364 pp.

——, B. A. Callander, and S. K. Varney, 1992: *Climate Change 1992. The Supplementary Report to the IPCC Scientific Assessment.* Cambridge University Press, 200 pp.

Hoyt, D. V., and K. H. Schatten, 1993: A discussion of plausible solar irradiance variations, 1700–1992. *J. Geophys. Res.,* **98,** 18 895–18 906.

Jones, P. D., 1994a: Recent warming in global temperature series. *Geophys. Res. Lett.,* **21,** 1149–1152.

——, 1994b: Hemispheric surface air temperature variations: A reanalysis and an update to 1993. *J. Climate.,* **7,** 1794–1802.

——, and K. R. Briffa, 1992: Global surface air temperature variations during the twentieth century: Part 1, spatial, temporal and seasonal details. *The Holocene,* **2,** 165–179.

——, S. C. B. Raper, R. S. Bradley, H. F. Diaz, P. M. Kelly, and T. M. L. Wigley, 1986a: Northern Hemisphere surface air temperature variations, 1851–1984. *J. Climate Appl. Meteor.,* **25,** 161–179.

——, ——, and T. M. L. Wigley, 1986b: Southern Hemisphere surface air temperature variations, 1854–1984. *J. Climate Appl. Meteor.,* **25,** 1213–1230.

——, T. M. L. Wigley, and G. Farmer, 1991: Marine and land temperature data sets: A comparison and a look at recent trends. *Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations,* M. E. Schlesinger, Ed., Elsevier Science, 593–602.

Karoly, D. J., J. A. Cohen, G. A. Meehl, J. F. B. Mitchell, A. H. Oort, R. J. Stouffer, and R. T. Wetherald, 1994: An example of fingerprint detection of greenhouse climate change. *Climate Dyn.,* **10,** 97–105.

Keeling, C. D., R. B. Bacastow, A. F. Carter, S. C. Piper, T. P. Whorf, M. Heimann, W. G. Mook, and H. Roeloffzen, 1989: A three-dimensional model of atmospheric CO_2 transport based on observed winds: 1) Analysis of observational data. *Aspects of Climate Variability in the Pacific and the Western Americas, Geophys. Monogr.,* No. 55, Amer. Geophys. Union, 165–236.

Lunkeit, F., R. Sausen, and J. M. Oberhuber, 1995: Climate simulations with the global coupled atmosphere–ocean model ECHAM2/OPYC. Part 1: Present-day climate and ENSO events. *Climate Dyn.,* **12,** 195–212.

Madden, R. A., and V. Ramanathan, 1980: Detecting climate change due to increasing carbon dioxide. *Science,* **209,** 763–768.

Maier-Reimer, E., and U. Mikolajewicz, 1989: Experiments with an OGCM on the cause of the Younger Dryas. *Oceanography 1988,* A. Ayala-Castanares and W. Wooster, Eds., UNAM Press, 87–100.

——, ——, and K. Hasselmann, 1993: Mean circulation of the Hamburg LSG model and its sensitivity to the thermohaline surface forcing. *J. Phys. Oceanogr.,* **23,** 731–757.

Manabe, S., and R. J. Stouffer, 1988: Two stable equilibria of a coupled ocean–atmosphere model. *J. Climate,* **1,** 841–866.

——, and ——, 1996: Low frequency variability of surface air temperature in a 1000 year integration of a coupled ocean–atmosphere model. *J. Climate,* **9,** 376–393.

Mitchell, J. F. B., T. J. Johns, J. M. Gregory, and S. F. B. Tett, 1995: Transient climate response to increasing sulphate aerosols and greenhouse gases. *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.

——, T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.,* **110,** 600–706.

——, K. Y. Kim, S. S. P. Shen, and J. W. Hardin, 1995: Detection of forced climate signals. Part I: Filter theory. *J. Climate,* **8,** 401–408.

Oberhuber, J. M., 1993a: Simulation of the Atlantic circulation with a coupled sea ice–mixed layer–isopycnical general circulation model. Part I: Model description. *J. Phys. Oceanogr.,* **23,** 808–829.

——, 1993b: Simulation of the Atlantic circulation with a coupled sea ice–mixed layer–isopycnical general circulation model. Part II: Model experiments. *J. Phys. Oceanogr.,* **23,** 830–845.

Parker, D. E., P. D. Jones, C. K. Folland, and A. Bevan, 1994: Interdecadal changes of surface temperature since the late nineteenth century. *J. Geophys. Res.,* **99,** 14 373–14 399.

Pennell, W. T., and Coauthors, 1993: The detection of anthropogenic climate change. *Proc. Fourth Symp. on Global Change Studies,* Anaheim, CA, Amer. Meteor. Soc., 21–28.

Roeckner, E., and Coauthors, 1992: Simulation of the present-day climate with the ECHAM model: Impact of model physics and resolution. MPI Rep. 93, 172 pp. [Available from Max-Planck-Institut für Meteorologie, Bundesstr. 55, 20146 Hamburg, Germany.]

Santer, B. D., T. M. L. Wigley, P. D. Jones, and M. E. Schlesinger, 1991: Multivariate methods for the detection of greenhouse-gas-induced climatic change. *Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations,* M. E. Schlesinger, Ed., Elsevier Science, 511–536.

——, U. Cubasch, U. Mikolajewicz, and G. C. Hegerl, 1993a: The use of general circulation models in detecting climate change induced by greenhouse gases. PCMDI Rep. 10, 30 pp. [Available from Lawrence Livermore National Laboratory, Livermore, CA 94550.]

——, T. M. L. Wigley, and P. D. Jones, 1993b: Correlation methods in fingerprint detection studies. *Climate Dyn.,* **8,** 265–276.

——, W. Brüggemann, U. Cubasch, K. Hasselmann, H. Höck, E. Maier-Reimer, and U. Mikolajewicz, 1994: Signal-to-noise analysis of time-dependent greenhouse warming experiments. Part 1: Pattern analysis. *Climate Dyn.,* **9,** 267–285.

——, K. E. Taylor, J. E. Penner, T. M. L. Wigley, U. Cubasch, and P. D. Jones, 1995a: Towards the detection and attribution of an anthropogenic effect on climate. *Climate Dyn.,* **12,** 77–100.

——, U. Mikolajewicz, W. Brüggemann, U. Cubasch, K. Hasselmann, H. Höck, E. Maier-Reimer, and T. M. L. Wigley, 1995b: Ocean variability and its influence on the detectability of greenhouse warming signals. *J. Geophys. Res.,* **100,** 10 693–10 725.

——, T. M. L. Wigley, T. P. Barnett, and E. Anyamba, 1996a: Detection of climate change and attribution of causes. *Climate Change 1995. The IPCC Second Scientific Assessment.* 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.

Sausen, R., K. Barthel, and K. Hasselmann, 1988: Coupled ocean–atmosphere models with flux correction. *Climate Dyn.,* **2,** 145–163.

Stouffer, R. J., S. Manabe, and K. Y. Vinnikov, 1994: Model assessment of the role of natural variability in recent global warming. *Nature,* **367,** 634–636.

von Storch, H., and E. Roeckner, 1983: Methods for the verification of general circulation models applied to the Hamburg University GCM. Part I: Test of individual states. *Mon. Wea. Rev.*, **111,** 1965–1976.

von Storch, J., 1994: Interdecadal variability in a global coupled model. *Tellus,* **46A,** 419–432.

——, and G. Hannoschöck, 1986: Statistical aspects of estimated principal vectors (EOFs) based on small sample sizes. *J. Climate Appl. Meteor.,* **24,** 716–724.

Tahvonen, O., H. von Storch, and J. von Storch, 1993: Economic efficiency of CO_2 reduction programs. MPI Rep. 105, 26 pp. [Available from Max-Planck-Institut für Meteorologie, Bundesstr. 55, 20146 Hamburg, Germany.]

Thiebaux, H. J., and F. W. Zwiers, 1984: The interpretation and estimation of effective sample size. *J. Climate Appl. Meteor.,* **23,** 800–811.

Wigley, T. M. L., and T. P. Barnett, 1990: Detection of the greenhouse effect in the observations. *Climate Change. The IPCC Scientific Assessment,* J. T. Houghton et al., Eds., Cambridge University Press, 239–255.

——, and B. D. Santer, 1990: Statistical comparison of spatial fields in model validation, perturbation and predictability experiments. *J. Geophys. Res.,* **95,** 851–865.

——, and S. C. B. Raper, 1990: Natural variability of the climate system and detection of the greenhouse effect. *Nature,* **344,** 324–327.

——, and ——, 1991: Internally generated natural variability of global mean temperature. *Greenhouse-Gas-Induced-Climatic Change: A Critical Appraisal of Simulations and Observations.* M. E. Schlesinger, Ed., Elsevier Science, 471–482.

Zwiers, F. W., and H. von Storch, 1995: Taking serial correlation into account in tests of the mean. *J. Climate,* **8,** 336–351.