

## Detecting Climate Signals Using Space–Time EOFs

GERALD R. NORTH AND QIGANG WU

*Department of Atmospheric Sciences, Texas A&M University, College Station, Texas*

(Manuscript received 15 October 1999, in final form 16 August 2000)

### ABSTRACT

Estimates of the amplitudes of the forced responses of the surface temperature field over the last century are provided by a signal processing scheme utilizing space–time empirical orthogonal functions for several combinations of station sites and record intervals taken from the last century. These century-long signal fingerprints come mainly from energy balance model calculations, which are shown to be very close to smoothed ensemble average runs from a coupled ocean–atmosphere model (Hadley Centre Model). The space–time lagged covariance matrices of natural variability come from 100-yr control runs from several well-known coupled ocean–atmosphere models as well as a 10 000-yr run from the stochastic energy balance climate model (EBCM). Evidence is found for robust, but weaker than expected signals from the greenhouse [amplitude  $\sim 65\%$  of that expected for a rather insensitive model (EBCM:  $\Delta T_{2\times\text{CO}_2} \approx 2.3^\circ\text{C}$ )], volcanic (also about 65% expected amplitude), and even the 11-yr component of the solar signal (a most probable value of about 2.0 times that expected). In the analysis the anthropogenic aerosol signal is weak and the null hypothesis for this signal can only be rejected in a few sampling configurations involving the last 50 yr of the record. During the last 50 yr the full strength value (1.0) also lies within the 90% confidence interval. Some amplitude estimation results based upon the (temporally smoothed) Hadley fingerprints are included and the results are indistinguishable from those based on the EBCM. In addition, a geometrical derivation of the multiple regression formula from the filter point of view is provided, which shows how the signals “not of interest” are removed from the data stream in the estimation process. The criteria for truncating the EOF sequence are somewhat different from earlier analyses in that the amount of the signal variance accounted for at a given level of truncation is explicitly taken into account.

### 1. Introduction

This paper provides estimates of the strengths of the linear responses in the surface temperature field to the four forcings: greenhouse gases (hereafter abbreviated  $G$ ), anthropogenic aerosols ( $A$ ), volcanic dust veils ( $V$ ), and the 11-yr solar cycle ( $S$ ). In so doing we also evaluate the 90% confidence regions for the amplitude estimates, individually and in pairs. The estimates make use of natural variability calculations from a variety of current coupled ocean–atmosphere general circulation models; these natural variability calculations go into a regression model for the signal strengths. The space–time signal patterns used in the regression come from our own energy balance climate model (and as a check from the Hadley Centre Model). The estimates are then extracted from a 100-yr data stream in several combinations of 36 and 72 stations that have good coverage over the last 50 and 100 yr.

Hasselmann (1979) was perhaps the first to consider such a technique in conjunction with this class of problems. In his first paper on the subject he proposed an optimal weighting scheme to examine the signals in the seasonal climate problem, but in later papers he has focused on the long-term climate problem (e.g., Hasselmann 1993; Hegerl et al. 1996). A significant contribution was made by Bell (1982, 1986) who first looked at the long-term problem and applied his optimal weighting formalism to the existing evidence available at that time. After a series of papers by Barnett and colleagues (e.g., Barnett 1986, 1991; Barnett and Schlesinger, 1987), a kind of prototype analysis has emerged (e.g., North et al. 1995; North and Kim 1995; Leroy 1998; Stevens and North 1996; North and Stevens 1998; this last paper will be referred to as NS98). A recent summary of activity in the field is given by Barnett et al. (1999) and a very recent report of results comes from Tett et al. (1999). The approach taken by North and colleagues (following Hasselmann) was to develop a “filter” through which to pass the data stream. Emerging from the filtered data stream are estimates of the strengths of the signal waveforms. The filter is to have an optimality property that maximizes the signal-to-noise characteristics of the estimate of the signal strength being sought.

---

*Corresponding author address:* Gerald R. North, Dept. of Meteorology, Texas A&M University, College Station, TX 77843-3150.  
E-mail: g-north@tamu.edu

Allen and Tett (1999; hereafter referred to as AT99) have stressed the fact that when we consider the multiple signal problem, we are simply performing the classic multiple regression analysis with the linear statistical model:

$$T^{\text{data}}(\mathbf{r}, t) = \sum_s \alpha_s S_s(\mathbf{r}, t) + N(\mathbf{r}, t), \quad (1)$$

where  $T^{\text{data}}(\mathbf{r}, t)$  represents the space,  $\mathbf{r}$ , and time,  $t$ , dependence in the stream of data,  $S_s(\mathbf{r}, t)$  is the  $s$ th signal whose shape in space–time must be specified in advance (the fingerprint),  $s = G, A, V, S$ , and  $N(\mathbf{r}, t)$  is the so-called natural variability. The strengths of the signals  $\alpha_s$  are to be determined by standard regression analysis. While the filter interpretation of Hasselmann and North and their collaborators is extremely useful from an intuitive point of view, the multiple regression interpretation is slightly less mysterious to the general scientific audience. In what follows we will use each interpretation as convenient. Some of the more technical comparisons of the two points of view are given in appendixes A and B.

Our approach makes use of space–time empirical orthogonal functions (EOFs) to represent all the fields involved. These EOFs may be defined as the eigenvectors of the covariance matrix coupling different space–time points:

$$\mathbf{K}(\mathbf{r}, t; \mathbf{r}', t') = \langle T(\mathbf{r}, t)T(\mathbf{r}', t') \rangle, \quad (2)$$

where  $\langle \cdot \rangle$  stands for ensemble average.

In practice the space is discretized so that we use a finite number of stations where data are thought to be well controlled for quality, etc. Also the time is usually discretized as monthly, seasonal, or annual averages (in this study we use annual averages). The length of the record is also important. In this study we use a variety of record lengths and choices of stations [20 tropical stations of 100-yr duration; 43 tropical stations, 20 with 100 yr, 23 with 50 yr; 36 stations globally distributed with 100 yr; and 72 stations, half with 100 yr and half with 50 yr; in one suite we use only (the last) 50 yr with all 72 stations]. The eigenvalue problem boils down to

$$\sum_{\mathbf{r}', t' \in \mathcal{D}} \mathbf{K}_{\mathbf{r}, t; \mathbf{r}', t'} \psi_n(\mathbf{r}', t') = \lambda_n \psi_n(\mathbf{r}, t), \quad (3)$$

where  $\psi_n(\mathbf{r}, t)$  is the  $n$ th eigenvector of  $\mathbf{K}$ , and  $\lambda_n$  is the associated eigenvalue that represents the variance associated with the  $n$ th mode. The domain  $\mathcal{D}$  is nominally the 100-yr interval (1894–1993) combined with 20, 36, or more station combinations. In some cases  $\mathcal{D}$  involves a combination of some stations with 100 yr and some with 50 yr. Note that the square matrix  $\mathbf{K}$  is of very large dimension that will require special considerations

in the analysis. Now all fields in the problem can be expanded into the complete set of space–time EOFs, for example,

$$T(\mathbf{r}, t) = \sum_n T_n \psi_n(\mathbf{r}, t) \quad (4)$$

and (1) becomes (in EOF mode form)

$$T_n^{\text{data}} = \sum_s \alpha_s S_{s,n} + N_n \quad (5)$$

with  $\langle N_n N_m \rangle = \lambda_n \delta_{nm}$ .

When multiple signals are involved, they are unlikely to be orthogonal to one another. Hence, components of unwanted signals will have components along the direction of the signal of interest. This potential interference effect is dealt with automatically in the normal course of the multiple regression analysis. In the filter formalism, the unwanted signals are removed by deleting all components of the data stream that are parallel to the unwanted signals (see appendix B).

Complicating the filter process are errors due to any incorrect specifications of any signal patterns not of interest. Erroneous specifications of signal patterns not of interest, whatever their origin, obviously lead to “leakage” and therefore a bias in the estimate of the amplitude of the signal of interest. We do not deal with this problem in the present study, but set the stage for it in future work.

Another important aspect of the multiple signal problem is that the (sample-dependent) random errors committed in estimating the various amplitudes are correlated (see the discussion by AT99). For example, an overestimate of  $G$  is expected to be accompanied by an overestimate of  $A$ , primarily because of the near anticollinearity of these two vectors (specified space–time patterns)—an erroneously strong greenhouse gas can be effectively cancelled by an erroneously strong anthropogenic aerosol response because the space–time fingerprints of these two influences are spatially and temporally so similar (but opposite in sign) and therefore difficult to distinguish. Angles between the various vectors as given by NS98 (for the particular “solar band” EOF subset they used) indicate near orthogonality for all except the  $A$ – $G$  pair. This will also be apparent in the error ellipses that have their principal axes along the  $x$  and  $y$  axes for all but the  $A$ – $G$  pair to appear later in our results.

Two major problems must be addressed early on by the investigator:

- 1) How to correctly characterize the signal waveforms,  $S_s(\mathbf{r}, t)$ .
- 2) How to compute the covariance matrix,  $\mathbf{K}$ , or its equivalent the  $\psi_n(\mathbf{r}, t)$  and the  $\lambda_n$ .

For the input signals (1) we use the Energy Balance Climate Model (EBCM) developed by Stevens in Stevens and North (1996), NS98, and Stevens (1997); there

Locations of 72 detection boxes

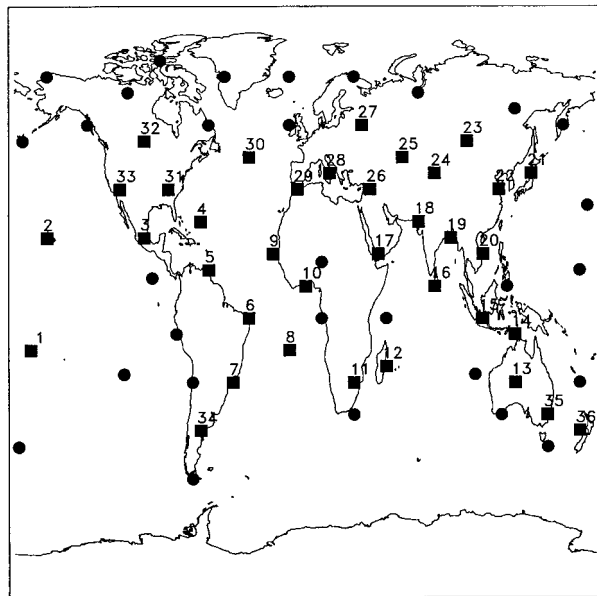


FIG. 1. Locations of the 72 detection boxes. Each of the 36 numbered boxes has 100 yr of observational data (1894–1993) from the Jones dataset; each of the 36 locations indicated by disk-shaped markers has 50 yr of data (1944–93).

were no changes in the model from these earlier reports. Simulations by the EBCM presumably are less realistic than those generated by current ocean–atmosphere coupled global climate models (however, see Figs. 1 and 2 for a direct comparison). As can be seen in Fig. 2, the EBCM has the advantage of carrying along no sampling errors associated with estimates of the signal based upon a small ensemble. That is to say, the signal is purely deterministic without any corruption added to it from natural variability. It is known that such noise corruption in the signal causes a bias as well as an enlarged random error in the estimation of the signal amplitudes (AT99; M. J. Stephens, G. R. North, and Q. Wu 2000, unpublished manuscript). In GCM studies of detection, this problem is dealt with by aggregating into decadal or longer averages or trends in order to reduce the noise corrupting the trial signal. The trade-off in this aggregation process is to sacrifice the ability to detect or properly account for the volcanic and solar cycle signals. In principle, all signals should be detected simultaneously. Because the space–time EOFs were established from very long control runs each component (actually a quadruple of components because there are four signals) represents a statistically independent estimator of the four signal amplitudes; all the estimators are ultimately optimally combined in the regression process to form the least mean square error estimate of the signal amplitudes (see appendix B). Of course, even 1 k control runs do not eliminate all sampling error, so that the sample EOFs are only asymptotically statistically in-

dependent. For a discussion of the differences between GCM generated signals see Hegerl et al. (2000).

In (2) we use 1000-yr control runs of several GCMs. We conduct these analyses from the available GCM runs separately and in parallel for comparison of final estimates of the  $\alpha_s$ , which we call  $\hat{\alpha}_s$  ( $\hat{\cdot}$  represents an estimate), at the end.

The present approach (following NS98) also poses a solution to the problem of where on the planet to conduct the estimation (of  $\alpha_s$ ) process. We choose specific station sites to evaluate both signal and data. These can be stations (points on the sphere) chosen well in advance where we know we have continuous, reliable (near) surface temperature data. Also consideration is given to choosing sites where we know the models are most reliable (e.g., not near sea–ice margins where model performance is suspect) for producing the signal waveform. In this way we do not smooth data, then smooth model output to a common format and compare (or, more precisely, process through the filter).

The use of space–time EOFs as a basis set has some unique advantages. As alluded to above, the EOF mode coefficients are uncorrelated (for sufficiently long control runs), which leads to a statistically independent estimator for each (foursome of) modes. This is to be contrasted with studies that examine the time series of various statistical indexes or detection variables. In our opinion the fully spectral EOF approach lends itself to a cleaner statistical characterization of the estimates of signal amplitudes and the associated uncertainties.

A major problem faced by every investigator working on this problem is how is the EOF sequence to be truncated? In our present approach there will be thousands of space–time EOFs. Experience from NS98 and others suggests that the large-index EOFs correspond to small  $N$  time- and space scales and are probably unreliable in the control simulations and probably also unreliable in the trial signal waveforms as well. Hence, it makes sense to truncate this sequence at some reasonable order to prevent absurd answers for the signal-to-noise ratio (SNR). Another consideration is the amount of the signal in question that is projected onto a particular EOF. In other words, we want to capture as much of the signal's shape as possible in the truncation agreement. An objective measure is given by the amount of variance of the total signal (variance summed over the space–time domain  $\mathcal{D}$ ) accounted for by a given truncation level (thanks to S. Leroy for this idea). Inevitably some subjectivity will creep into the decision process at this point. We indicate our motivation for the choice of truncation level by graphing some of the performance indicators as a function of truncation level.

We wish to make one final point about the solar signal. We follow NS98 in not using any long timescale changes in the solar signal partly because we have very little faith in any time dependence that might be implemented. Instead we concentrate on the 11-yr component only, since the amplitude of this forcing is well established

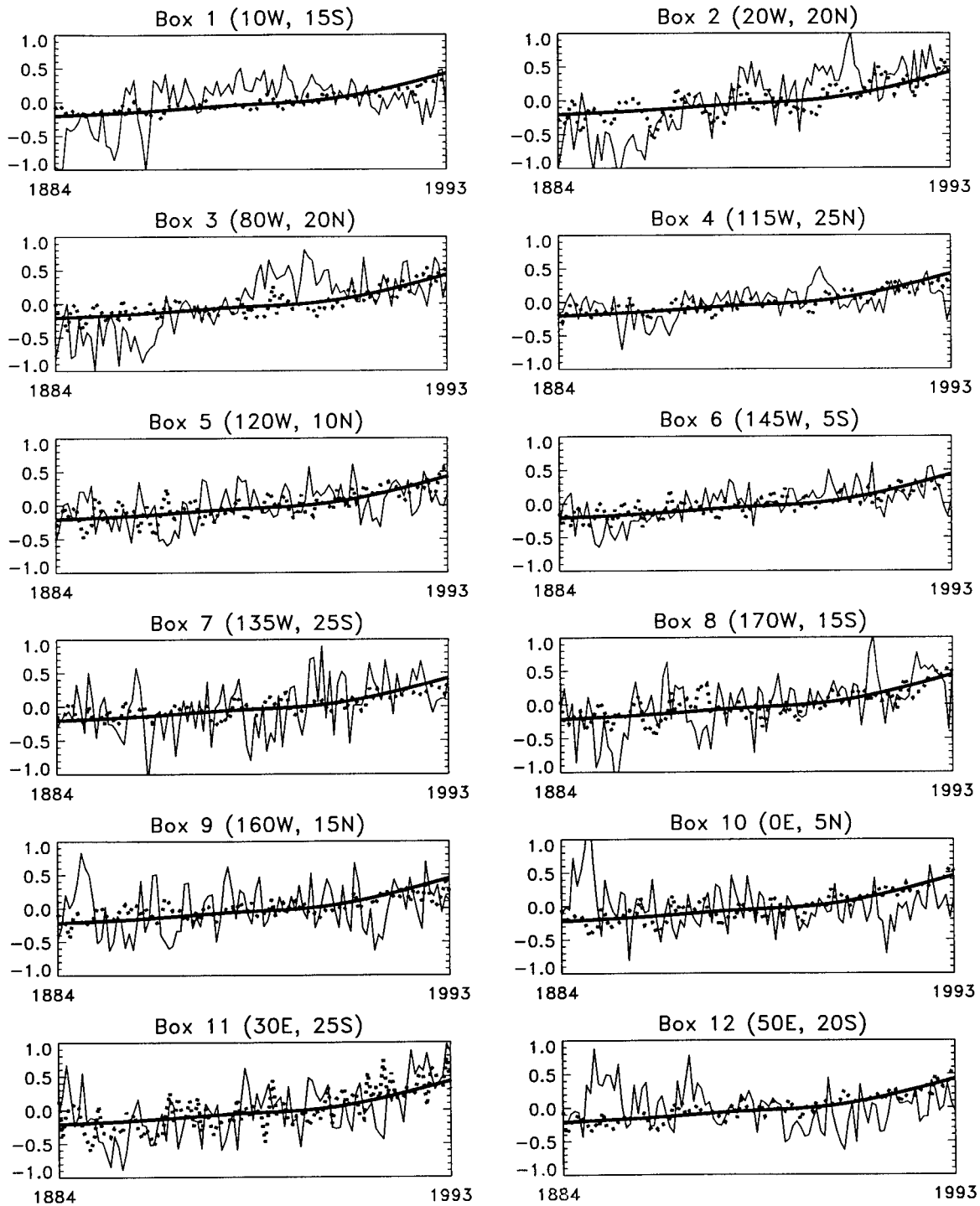


FIG. 2. Each panel shows modeled and observed time series from a different observational site as indicated in Fig. 1. Greenhouse gas signal from the EBCM (thick solid line) multiplied by 0.65 (in conformity with our detection results). The dotted line is an average across a four-member ensemble of HadCM2 forced by greenhouse gases (also multiplied by 0.65 to conform with our detection results). Observational data from Jones are shown by the thin solid line.

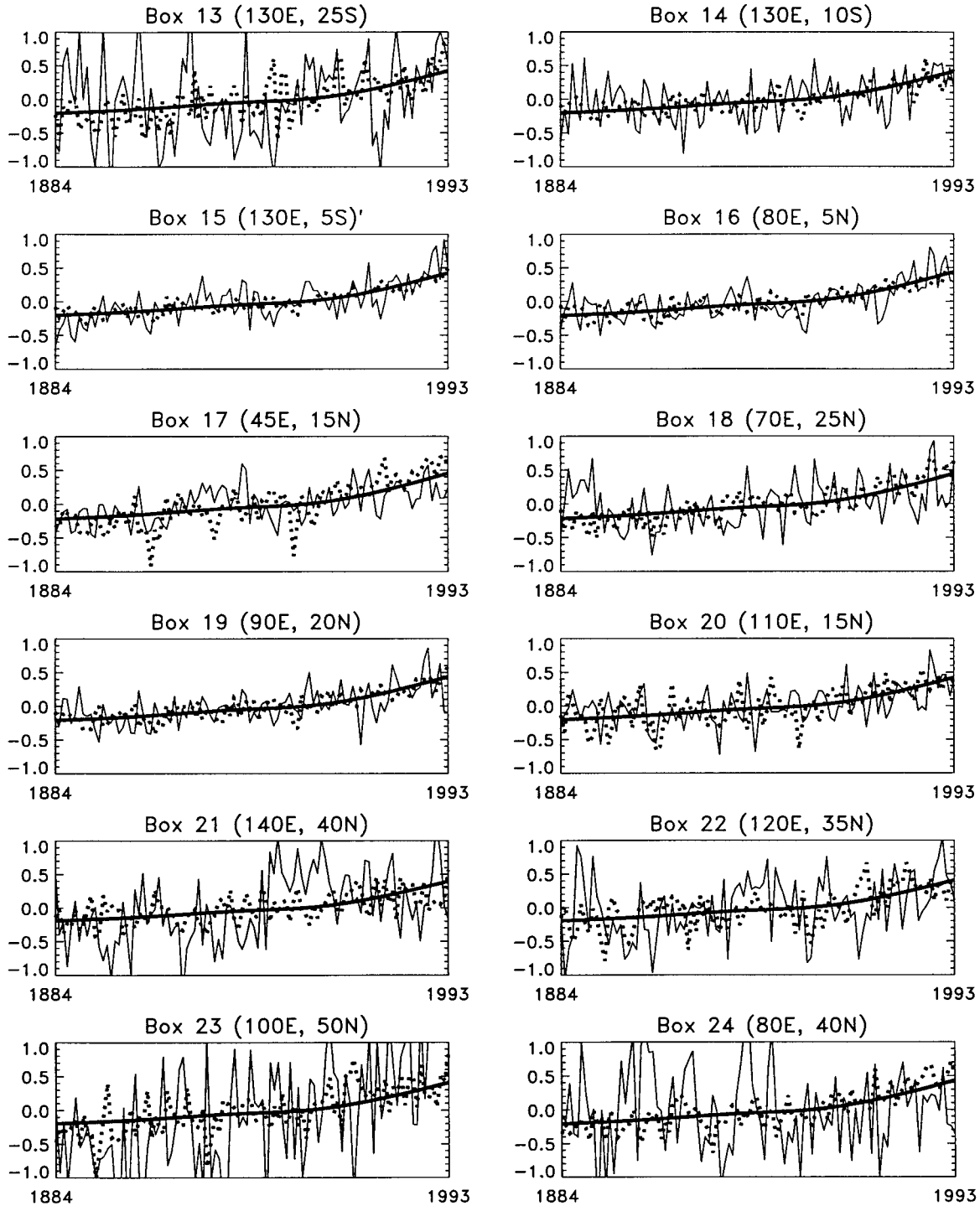


FIG. 2. (Continued)

by satellite measurement and it is straightforward to model. We emphasize that the detection of this component is of purely scientific interest—we do not consider the solar cycle to be an important part of climate change from a practical point of view. On the other hand, it is important for the theory of climate since it is one

of the few external forcings at the decadal scale that can be used to test climate models. To some extent the same holds for volcanic signals as emphasized by NS98. In both of these cases it is important to maintain a rather high resolution in time in the detection process (see the spectral components as displayed in NS98).

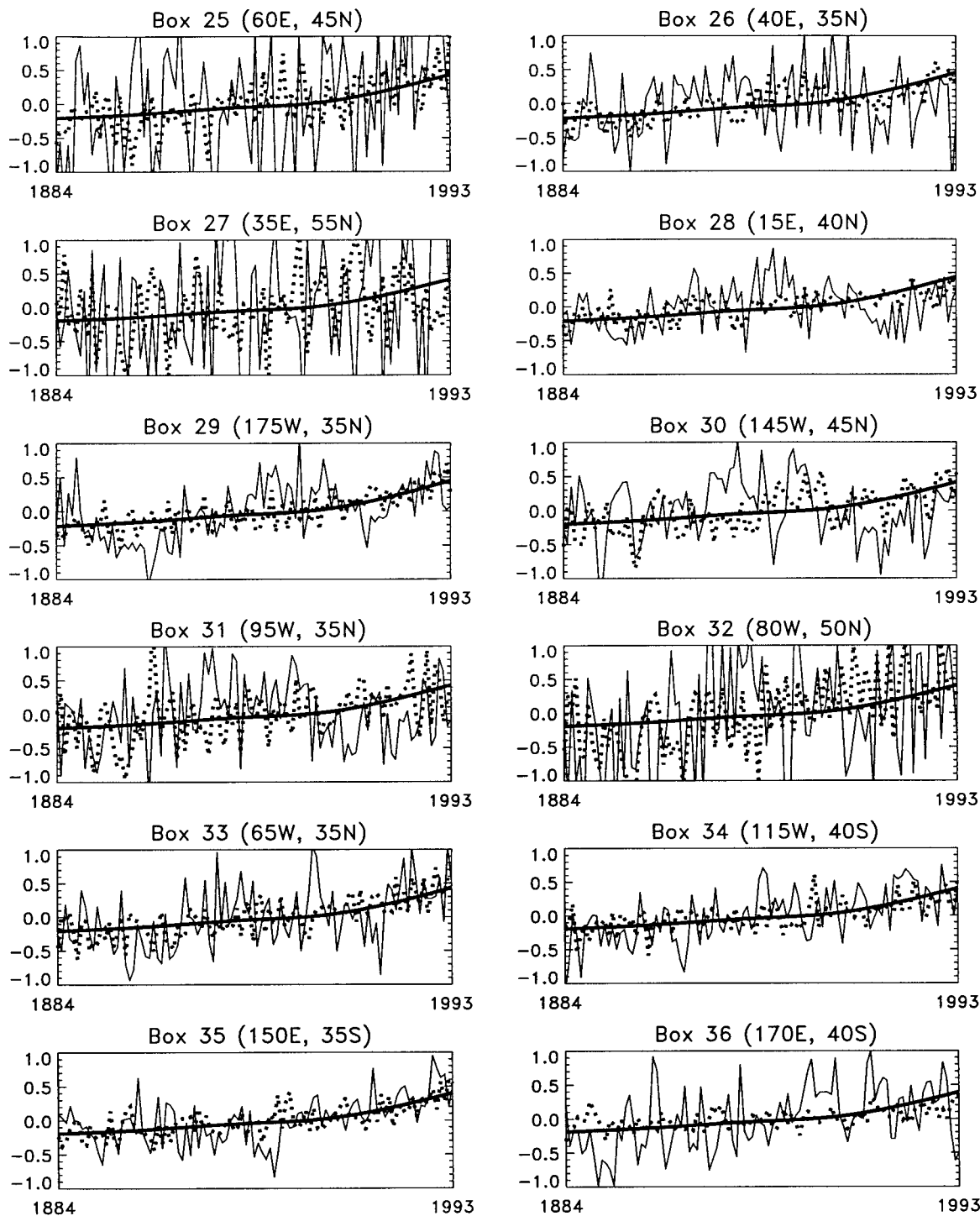


FIG. 2. (Continued)

## 2. Approach and comparison to earlier work

### a. NS98 comparison

The aims of this paper are similar to the earlier paper NS98, but in the present paper we remove or improve upon several important approximations that might have

influenced the results reported there. These differences are reviewed next.

- 1) In NS98 the time series of annual entries was assumed to be stationary and it was further assumed that the 100-yr record used was sufficiently long that

the eigenvectors from an infinitely long segment could be used. This means that the temporal part of the EOFs could be factored out as Fourier components (see, e.g., North 1984). In other words, each frequency component had attached a distinct string of spatial EOFs. Since in NS98 only a narrow band from frequencies  $20 \text{ (yr)}^{-1}$  to  $7 \text{ (yr)}^{-1}$  (referred to as the solar band), was used, this might have been expected to be a good approximation. However, experience has shown that this approximation has a tendency to overestimate the signal-to-noise ratio for the estimation of the signal amplitude, especially if lower frequencies are included. Therefore, in the present work we have found the exact space–time eigenvectors for the 100-yr interval ( $\equiv \mathcal{D}$ ; in some cases 50 yr and in some others a mix of 50- and 100-yr intervals). In this analysis we do not exclude any band of frequencies a priori, but note that higher frequencies contribute very little to the estimation performance because the signals have very little projection on these modes. While the Fourier frequency analysis was convenient computationally and conceptually, there is really no reason to avoid the exact treatment. Of course, one has to ask about the adequacy of 1000-yr runs for estimating the space–time EOFs. We have used one 10 k run with our stochastic EBCM to confirm that this is not a problem (appendix E). We also have intercompared (see tables later) the results from the different control runs and have found little difference in our estimates of signal amplitudes.

- 2) In NS98 the treatment of multiple signals was sub-optimal. In that work, in order to find the amplitude of signal  $S_s$ , the component of the signal vector perpendicular to the sum of the other three was considered in the data stream (allowed to pass through the filter; all other components were annihilated). This made the part of the signal of interest perpendicular to the other three dependent on their actual amplitudes. In the present work we use the standard multiple regression formalism that has the geometrical interpretation of filtering out the unwanted components simultaneously (see appendix B) without regard to their amplitudes.
- 3) In the present work we consider the possibility of records of unequal length. We work the problem end to end with 20 tropical stations with 100-yr records; with 20 tropical 100-yr records and 23 50-yr records together; and finally with 72 stations spread over the globe, half with 100, half with 50-yr records. We also present a case with 72 stations and (the last) 50 yr of data for each. (See Fig. 1.)
- 4) In the present work, account is taken of the correlation of random errors in  $\hat{\alpha}_s$  with  $\hat{\alpha}_{s'}$ ;  $s \neq s'$ . This leads to error ellipsoids that can be viewed to see if the zero plane of any signal strength slices the ellipsoidal confidence volume. (If so, the signal has

questionable significance; i.e., the null hypothesis cannot be rejected at the 10% level.)

- 5) In the present analysis we used all 12 months to form annual averages in all fields. This contrasts with NS98 where only the summer half year was used at the midlatitude sites. This less arbitrary choice did not lead to an appreciable loss of SNR.

Many of the same assumptions apply to the present work as applied to NS98.

- 1) The basic linear superposition assumption implied by Eq. (1).
- 2) That the EBCM developed by Stevens is adequate to produce the signal waveforms (cf. Fig. 2).
- 3) That 1000-yr control runs of the GCMs are sufficient for the statistics that go into the space–time covariance matrix. Exactly as in NS98 we use control runs from the Max Planck Institute climate model (ECHAM1/LSG), two different versions of the Geophysical Fluid Dynamics Laboratory model (mixed layer ocean vs fully coupled deep ocean) and finally the most recent Hadley Centre Model (HadCM2); (Mitchell et al. 1995; Johns et al. 1997; Tett et al. 1996). We also use our stochastic EBCM (mixed layer ocean only) in the comparison; in this case we have a 10 k run that is broken into 1 k segments or in some cases the whole record is used to eliminate sampling error.

#### b. Comparison to Hadley Centre studies

Since our work is most similar to that contained in several papers coming out of the Hadley Centre (AT99, Tett et al. 1999; Stott et al. 2001, hereafter STJAIM), we give a very brief summary of their work as it relates to the present paper.

- 1) The Hadley group rely upon their GCM control runs and upon ensemble averages of several realizations of forced runs for their signal waveforms. By way of contrast, we use 1000-yr control runs from several different GCM runs and compare final results. We use our EBCM to generate all four signal waveforms. We feel based upon comparisons (shown later and in appendix C) that our signals and those from the Hadley GCM are not appreciably different.
- 2) Because of the limited lengths of available control runs and the sampling error associated with the estimation of the covariance matrix they have restricted their tests to 50-yr intervals. In forming their error ellipses or confidence intervals they have used separated segments of control runs to estimate the eigenvalues versus the error ellipses and confidence intervals. The argument for this is that in such a short sample, low-frequency variation might be very different from one sample to another and their procedure is reckoned to be more conservative. We tested this with a 10 k run of our mixed layer EBCM,

by using different 1 k segments for the eigenvalues and the ellipses. We found no practical differences in the sizes or shapes of the ellipses compared to other errors in the problem—some of these results are shown in appendix E.

- 3) In their analysis a spatial smoothing is performed on the data and the model output for variability and signals. For example, only scales of spherical harmonic degree 4 or lower are retained. When no data are available zeros are substituted for anomalies, that is, a mask is applied to both data and model signals. Similarly a temporal smoothing is applied in the form of decadal averages. For the 50-yr segment then there are only five temporal entries along with 25 spatial modes. Of the 125 possible space–time modes only 12–15 modes are typically retained in their analyses.

The temporal smoothing essentially precludes the estimation of the volcanic signal amplitude that has most of its power in the decadal range (see NS98 for a power spectrum). In our treatment, we avoid the smoothing altogether by choosing a priori 36 stations (each actually an average from 4 nearby stations within a  $10 \times 10$  degree box) of length 100 yr. We also use the last 50-yr segment from 72 stations. This is our “mask.” A possible difference is that in our 72 stations we purposely avoided polar stations where we thought models might be unreliable.

The number of space–time EOF modes in the two studies is vastly different because in the Hadley case the smoothing is done prior to the statistical estimation procedure. In our case we do no prior smoothing, leaving us with a large number of space–time EOFs. Our aggregation of data comes at the end as all the EOFs are used in the estimation problem. The benefits of data averaging come at the end rather than at the beginning. We compare these approaches in appendix G.

- 4) Our time series for detection ends in 1993 compared to 1996 in the Hadley studies. We repeated some of our calculations for the same period and found no differences.

### 3. Notation

Before presenting results some notation and details of the concepts will be listed. Based upon standard multiple regression analysis, the optimal estimator for a particular  $\alpha_s$  is given by

$$\hat{\alpha}_s = \sum_{s'} [(S^T \mathbf{W} \cdot S)^{-1}]_{s,s'} S_s^T \mathbf{W} T^{\text{data}}, \quad (6)$$

where the hat indicates that the variable is a statistical estimator, that is, a random variable dependent on the data sample. This random variable acquires its randomness from that in the data vector on the rhs, which con-

tains only one realization. A different realization of data would lead to a different realization of  $\hat{\alpha}_s$ . The implied (suppressed) matrix indices are over the EOF mode numbers, whereas the explicitly indicated indices are over the signals  $s, s' = (G, A, V, S)$ . We use bold face when the EOF mode index is suppressed. The matrix  $\mathbf{W}_{mn} \equiv \lambda_n^{-1} \delta_{mn}$  is the inverse space–time-lagged covariance matrix of the natural variability in its EOF or diagonal form,  $\mathbf{W} = \mathbf{K}^{-1}$ ; it forms a metric tensor in the space (Hasselmann 1993; AT99). An alternative derivation in which this metric tensor comes out naturally is given in Appendix A using the maximum likelihood method. Note that the indices  $s, s'$ , etc., denote signals, while indices  $n, m$ , etc., denote EOF mode number.

When the estimator  $\hat{\alpha}_s$  is written as in the last equation, the data stream is seen as being multiplied by an operator or a filter. Also the estimate of  $\alpha_s$  can be viewed as a sum of contributions from EOF components. Each four member subset (because of four signals, see Appendix B) of this sum represents a statistically independent estimator of  $\alpha_s$  (of course, if the series is truncated, the normalization has to be recalculated to restore zero bias). Also we note that if the EOFs are calculated from a finite length record, there will be some correlation between EOF amplitudes thereby reducing the number of degrees of freedom.

In the single signal case this takes the form:

$$\hat{\alpha} = \frac{1}{\gamma^2} \sum_m \frac{S_m}{\lambda_m} T_m^{\text{data}}, \quad (7)$$

where

$$\gamma^2 = \sum_m \frac{S_m^2}{\lambda_m}. \quad (8)$$

This latter has the interpretation of being a theoretical (no data involved) signal-to-noise ratio squared, and this decomposition according to EOF mode is useful in deciding on a level of truncation. For example, as the number of terms in the series increases, the values of the  $\lambda_m$  decrease toward zero, potentially giving a spuriously large SNR<sup>2</sup>. Often one can study the value of  $\gamma^2$  as a function of truncation level and make a sensible decision on where the truncation should occur.

It must be kept in mind that in practice one only has sample EOFs. The last two formulas hold if the EOFs are the exact ones for the population (Karhunen–Loève Functions). Furthermore, it is assumed that the model of signals plus noise is correct. For EOFs computed from finite samples there is a high bias in the estimated eigenvalues for low-index EOF components and a low bias for very high-index components. We tested this with different 1 k yr segments of our 10 k yr control run of the noise-forced EBCM. We found that the low-index eigenvalues derived from the 1 k runs were as much as 30% above those derived from the 10 k yr run. This bias in the lowest-index eigenvalues will lead to an underestimate for the signal-to-noise ratio when only



a few terms are retained. We stress then that (6) and (7) are mainly useful as guides in the truncation decision. The same bias effect will lead to somewhat larger confidence regions than if the eigenvalues were exact. On the other hand, the bias in eigenvalue estimation does not introduce a bias in the amplitude estimation; rather, it renders the estimate suboptimal by using biased weights.

A similar role is played by a matrix in the multiple signal problem:

$$\gamma_{ss'}^2 = \sum_n \frac{S_{sn} S_{s'n}}{\lambda_n} = \mathbf{SWS}' \quad (9)$$

The matrix  $\mathbf{\Gamma}$  can be formed as the array of the  $\gamma_{ss'}^2$ :

$$\mathbf{\Gamma}_{ss'} = \gamma_{ss'}^2 \quad (10)$$

Then the covariance matrix of the estimators  $\hat{\alpha}_s$  and  $\hat{\alpha}_{s'}$  is just

$$\text{cov}(\alpha_s, \alpha_{s'}) = (\mathbf{\Gamma}^{-1})_{ss'} \quad (11)$$

The quadratic form generated by these matrix elements set equal to a constant defines an ellipsoid centered at  $\langle \hat{\alpha}_1 \rangle, \dots, \langle \hat{\alpha}_4 \rangle$  in the four-dimensional space. If the constant is adjusted to the proper value one can generate the ellipsoid within which the point  $(\hat{\alpha}_1, \dots, \hat{\alpha}_4)$  should lie with 90% probability. This will be called the 90% confidence volume. We can study subdimensional ellipsoids by setting some of the  $\alpha_s$  variables to their nominal values (unity). This is equivalent to ignoring these variables in the regression analysis. We will be especially interested in pairs of the variables so that the confidence volumes can be shown as ellipses in a two-dimensional plane. If we set three of the variables to their nominal values we have simply a confidence interval. Then we can use  $\gamma_{ss}^2$  as an effective signal-to-noise ratio squared. This is the meaning of the symbol  $\gamma$  in the tables and figures.

#### a. Data and locations

Figure 1 shows 72 locations on the planet at which data are used in this study. The 36 numbered square boxes indicate the sites used by North and Stevens (1998). In each of these  $10 \times 10$  degree boxes, at least four continuous records exist for the period 1894–1993 and these time series were averaged across to form a single time series of one entry per year of annual averages. The circular boxes have been added for the present study. These sites have records covering the 50-yr period 1944–93.

All data streams are from the updated dataset archived and described by Jones and Briffa (1992); updated versions are available from the authors.

#### b. Signals

All signal waveforms ( $G, A, V, S$ ) were taken from the (unmodified) Stevens EBCM described in NS98.

Figures 2a–c show  $0.65 \times$  the EBCM-produced greenhouse signal (heavy solid line) at each of the 36 stations (square boxes) of Fig. 1 over 100 yr. The factor 0.65 reflects our optimal detection finding to be shown later. In each panel a four-member ensemble average of the same signal taken from the Hadley Centre GCM (lighter solid line) along with the observational data (dotted line) is shown. We also graphed the two-member signal from the ECHAM4 (Roeckner et al. 1999) forced model for the 36 stations (not shown, see the appendixes). The agreement (when temporally smoothed) was about as good as for the Hadley model (forced and control runs of various models are available through the Intergovernmental Panel on Climate Change). It is obvious that the GCM ensemble averages still contain considerable corruption from climate noise even after averaging two and four ensemble members. But one can also see that the EBCM does a credible job in capturing the rather weak geographical dependence of the greenhouse signal at least as given by the Hadley Centre model. As noted earlier, investigators using GCM signals have used decadal averaging or decadal trends to further reduce the error in their fingerprint patterns. Note that each model exhibits a rather good fit to the data even without invoking the other three signals.

#### c. Natural variability

We have used exactly the same natural variability estimates as described in NS98; namely, the space–time-lagged covariance matrices were computed from 1000-yr control runs of the GFDL mixed layer model (called here GFDLml) and the coupled ocean–atmosphere model of GFDL (called here GFDLc); we also used a 1000-yr control run from the coupled ocean–atmosphere Max Planck Institute Model (ECHAM1/LSG called here MPIgcm). We have added to the NS98 list by including estimates based on the natural variability of the Hadley Centre Group (HadCM2).

It is useful to recall that from the optimal filtering point of view, it can be proven that the choice of natural variability model does not directly bias the results (e.g., see the derivation of optimal weighting in NS98). Using the very best natural variability model can reduce the spread of the estimates  $\hat{\alpha}_s$ , but no bias is introduced.

## 4. Results

In this section we present the results of the several studies we have conducted.

#### a. Signal strength estimates—EBCM signals

Table 1 gives the results of our analysis for all four signal strengths ( $\alpha_s$ ) and theoretical SNR ( $\gamma_{ss}$ ) for each choice of natural variability model (GFDLc, GFDLml, EBM, ECHAM1/LSG = MPIgcm, HadCM2) for the 20 tropical stations. The meaning of  $\gamma$  in this table is the

TABLE 1. Estimations of signal-to-noise ratio ( $\gamma$ ) and strength ( $\alpha$ ) of four signals based on 20 tropical boxes (100-yr observational data) given five different model prescriptions for the natural variability: GFDLc, GFDLml, EBCM, MPI and HadCM2.

Signal	GFDLc		GFDLml		EBCM		ECHAM1/LSG		HadCM2	
	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$
<i>S</i>	0.59	1.07	0.74	1.94	0.79	1.89	0.75	1.86	0.50	2.81
<i>G</i>	3.00	0.74	3.92	0.84	3.59	0.61	4.11	0.59	2.56	0.68
<i>V</i>	2.64	0.64	3.04	1.05	2.49	0.71	3.27	0.81	2.10	0.82
<i>A</i>	0.86	0.30	1.11	0.78	1.26	-0.15	1.44	0.20	0.71	0.05

value of the SNR based upon setting the other three signal strength variables to their nominal values (unity) so that the confidence volume degenerates to a confidence interval. It is called a theoretical SNR because it really is not based upon the data but only on the input signals and the prescribed natural variability. Tables 2–5 give similar results for the other site—record configurations.

Noteworthy in all five tables is the rather unambiguous finding that  $\hat{\alpha}_G$  is significant ( $3.0 \leq \gamma_G \leq 6.08$ ), but with a somewhat smaller amplitude than the input signal ( $0.50 \leq \alpha_G \leq 0.84$ ). Keep in mind that the EBCM we used for the signal has a rather low sensitivity ( $\Delta T_{\Sigma \times \text{CO}_2}^{\text{global}} \sim 2.0^\circ\text{C}$ ). Also keep in mind that the EBCM used had only a mixed layer ocean, which means that it responds rather quickly ( $\sim 1$  yr) to external changes. The shallow ocean also leads to a less distinct land–sea pattern in the response to long-term forcing. Both sensitivity and the ocean treatment affect signal strength. On the other hand, the EBCM has time response and amplitude characteristics similar to the four-member ensemble averages of current GCMs as seen in Fig. 2 at least for the 100-yr interval. On a longer interval into the future, we expect these two forced model runs would differ because of their different sensitivities and their treatments of the oceans.

As in NS98 we also find that volcanic influences are very significant ( $2.49 \leq \gamma_V \leq 6.05$ ). Again as in the case of *G* we find a lower value of the signal strength ( $0.46 \leq \alpha_V \leq 0.71$ ).

A curious feature is that anthropogenic aerosols *A* are not found to be significant in most of our site/record/ (natural variability model) configurations. The signal strength is weak—typically much less than 0.5—and the best SNR is only of the order of 1.5. A notable exception to this is shown in Table 5 in which only the last 50 yr of data are used. In each configuration  $\hat{\alpha}_A$  is

large and in some cases well exceeding unity. In this case the large values of  $\hat{\alpha}_A$  seem to be coming at the expense of  $\hat{\alpha}_V$ .

The solar signal *S* is stronger than reported by NS98 but still the best SNR is around 1.58 with an associated amplitude of 1.90. The results of Table 3 suggest that we may be near detecting the solar signal, since this SNR corresponds to a (one sided) confidence value of 94%.

The results of the tables are also summarized in Fig. 3. In this representation it is clear how independent the values of  $\hat{\alpha}_s$  are of the noise model chosen. The error bars indicate the 90% confidence range. As the eye traverses across the panels of Fig. 3 from left to right one can form a kind of superensemble average of the signal amplitude estimates. The qualitative variance estimated from this procedure should give some indication of the robustness of our results.

It is interesting that as we progress from Table 1 to Table 5, increasing the number of stations and the record lengths, that the SNR tends to increase slightly as expected, but the increases are not dramatic. It appears from this study that increasing from the 36 stations adds 20%–30% to the SNR. Kim and Wu (2000) have shown in a single signal study that the SNR can be increased by  $\sim 15\%$  by including the seasonal cycle through a cyclostationary EOF approach. It appears that this is a promising way to increase performance.

### b. Error ellipses

In this section we describe some typical two-dimensional error ellipses that describe the 90% confidence regions for the simultaneous estimation of the amplitude pair shown on the coordinate axes. Figure 4 shows nine such ellipses that are typical of our findings. The upper row Figs. 4a, 4b, and 4c show ellipses for 72 stations,

TABLE 2. Same as Table 1 except for 36 global boxes (100-yr observational data).

Signal	GFDLc		GFDLml		EBCM		ECHAM1/LSG		HadCM2	
	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$
<i>S</i>	0.67	1.37	0.92	2.55	0.66	1.45	0.90	1.39	0.64	1.04
<i>G</i>	3.50	0.72	4.90	0.66	4.30	0.62	5.20	0.57	4.62	0.71
<i>V</i>	2.99	0.72	3.84	0.75	2.89	0.60	3.92	0.73	3.04	0.66
<i>A</i>	1.42	0.05	1.96	-0.36	1.70	-0.05	2.02	-0.37	1.92	0.02

TABLE 3. Same as Table 1 except for 43 tropical boxes (20 boxes have 100-yr data and 23 boxes have 50-yr data).

Signal	GFDLc		GFDLml		EBCM		ECHAM1/LSG		HadCM2	
	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$
<i>S</i>	0.87	2.02	0.93	1.94	0.85	1.76	1.12	1.88	0.64	2.83
<i>G</i>	4.66	0.76	4.41	0.67	3.88	0.62	6.74	0.56	3.38	0.76
<i>V</i>	3.87	0.80	3.58	0.85	3.56	0.61	4.67	0.50	2.81	0.86
<i>A</i>	1.26	0.85	1.23	0.43	1.32	-0.21	1.87	0.03	0.99	0.45

36 of which have 100 yr of data and 36 have only the last 50 yr of data (see Fig. 1). The second row Figs. 4d, 4e, and 4f shows the same but for all 72 sites with only the last 50 yr of data. The bottom row Figs. 4g, 4h, and 4i are for signals *G* and *GA* from the HadCM2 four-member ensemble average and a the *V* and *S* signals come from the EBCM; the site/record configuration is for all 72 sites and for the last 50 yr. The different ellipses in an individual panel represent the different choices of control run used to calculate the EOFs and eigenvalues.

In Fig. 4a we see the estimates  $\hat{\alpha}_G$  versus  $\hat{\alpha}_A$  (denoted on the graph simply *G* vs *A* for simplicity). As expected there is a strong correlation between these two estimators because of their near anticollinearity. Note that in all six of the panels in which *G* appears it is easily significant at the 10% level; this is to say the liner  $G = 0$  does not intersect any ellipse. This appears to be a very robust result across all natural variability models and site/record configurations. It is also noteworthy that the most probable value of *G* (centroid of the ellipses) lies in the neighborhood of 0.60. A similar result holds for the *V* amplitude: highly significant, but of most probable amplitude well below unity (in fact, the unity hypothesis would be rejected for both *G* and *V* in these panels).

In the case of the solar signal we turn especially to panel *c* that utilizes 100 yr of data ( $\sim 9$  cycles). It appears that when estimated simultaneously with the *A* signal it is significant at the 10% level. This is found to be true in combinations of *S* with the other variables as well. We conclude that the solar cycle amplitude is marginally significant at the 10% level and that its most probable value is in the range 1.0–2.5.

The aerosol signal *A* is more problematic. In panels *a* and *c* we cannot reject the null hypothesis ( $A = 0$ ). But when we look only at the last 50 yr of data, we find significance in Figs. 4d and 4f for two of the five

natural variability choices and one of the five is marginal. Note that it is the HadCM2 and ECHAM1/LSG models that indicate significant aerosol signals. The most probable value of the aerosol amplitude over the five ellipses in Figs. 4d or 4f is about 0.5 with the value of 1.0 present in about half the ellipses. The emergence of potentially significant *A* in the last 50 yr may be a physical occurrence because of its larger and possibly more readily discriminated signal during this period. This latter result is in harmony with the Hadley Centre reports (AT99; Tett et al. 1999, STJAIM). In the mode analysis for the *A* with only the last 50 yr of data, we found the truncation decision to be rather difficult. If we cut off at about mode 200 we obtain the very large value of  $\hat{\alpha}_A$ , but the value of  $\gamma^2$  is barely unity. If we enlarge the number of modes retained, the value of  $\hat{\alpha}_A$  becomes progressively smaller and well below unity, but in this range the eigenvalues become so small that this latter result is suspicious. Hence, we report the value of 1.42 in Table 5, but the corresponding value of  $\gamma$  is only 0.92. We note in passing that we also did a case where the last 50 yr ended in 1996 (as in the Hadley Centre work) and found the results to be indistinguishable from what is reported above.

Figs. 4g, 4h, and 4i. are discussed in the next section.

### c. Hadley Centre signals

We had available four-member ensemble averages from two forced runs with the HadCM2—a greenhouse-only signal (*G*) and a superposition of *G* and *A* that we call *GA* (note that in the Hadley papers the notation *GS* [*S* for sulfate aerosol] is used for what we call *GA*). We ran these cases to estimate  $\alpha_G$  and  $\alpha_{GA}$ . In order to smooth out the noise from these input signal waveforms we ran an 11-yr moving average of these fields. We inserted our EBCM signals for *V* and *S*, which had little

TABLE 4. Same as Table 1 except for 72 global boxes (36 boxes have 100-yr data and 36 boxes have 50-yr data).

Signal	GFDLc		GFDLml		EBCM		ECHAM1/LSG		HadCM2	
	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$
<i>S</i>	0.91	1.64	0.95	2.43	0.96	1.37	0.92	1.90	0.77	1.79
<i>G</i>	5.49	0.58	5.39	0.77	5.54	0.75	6.34	0.69	5.53	0.72
<i>V</i>	4.00	0.76	3.91	0.72	3.79	0.69	4.09	0.69	3.65	0.67
<i>A</i>	2.11	-0.30	2.23	-0.10	2.31	0.19	2.24	0.17	2.32	0.12

TABLE 5. Same as Table 1 except for 50-yr data (1944–93) for 72 detection boxes.

Signal	GFDLc		GFDLml		EBCM		ECHAM1/LSG		HadCM2	
	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$
<i>S</i>	0.70	1.40	0.56	1.76	0.68	1.70	0.52	3.28	0.52	1.48
<i>G</i>	3.66	0.72	3.53	0.92	3.88	0.64	3.78	0.53	3.11	0.70
<i>V</i>	3.00	0.40	2.35	0.57	2.86	0.35	2.50	0.40	2.43	0.50
<i>A</i>	1.03	0.85	0.96	1.28	1.12	0.75	1.05	0.84	0.92	1.42

effect because of the near orthogonality of these signals to *G* and *A*. The statistical model of the data was then

$$T_n^{\text{data}} = \alpha_G S_{Gn} + \alpha_{GA} S_{GAN} + \alpha_V S_{Vn} + \alpha_S S_{Sn} + N_n. \quad (12)$$

The true estimator of *G* is then  $\alpha_G + \alpha_{GA}$  and the true estimator of *A* is  $\alpha_{GA}$ . For comparison we show in Table 6 the results of two cases in which the HadCM2 signal pattern is used for all the different noise models. In this case we include only the signal patterns *G* and *GA* that are available from the Hadley Centre. For both *G* and *GA* there are four realizations for the ensemble average. In each case we use an 11-yr moving average to smooth the signal patterns. We show two cases one for the 72 station (combination of 100 yr and 50 yr as above) and for the 72 stations with only the last 50 yr of data. As discussed elsewhere in the paper, it appears that the *GA* is significant at the 10% level for the last 50 yr of data but questionable in the more extended dataset. Not shown in the table are the estimates of *V* and *S* because they were consistent with the earlier findings. As in the earlier results with the EBCM signals we find for *G*: 0.79, 0.61, 0.65, 0.62, and 0.54 for the combination 72/36 station 50/100 yr record, and 0.65, 0.87, 0.78, 0.75, and 0.79 for the 72 station/50-yr record; and for *A*: 0.10, 0.15, 0.26, 0.10, 0.19 (72/36 station 50/100 yr); 0.35, 0.49, 0.53, 0.41, and 0.44 (72 station, 100 yr).

Figures 4g, 4h and 4i show 90% confidence error ellipses for these signals based on 72 stations and the last 50 yr of data, with *G* and *GA* signals generated by the HadCM2 (four-member ensemble average, 11-yr moving averaged) while *V* and *S* signals come from the EBCM. Note that *G* is significant at the 10% level in all cases shown (which are typical) and that they are consistent with the findings of Figs. 4d, 4e and 4f, name-

ly that *A* and *G* are significant, but their most probable values are well below unity.

#### d. Analysis of mode contributions

In this section we show in some detail how the estimates depend on our choice of space–time EOF truncation level. The analysis will show the interplay between the several constraints in making this choice. In the analysis we graph (Fig. 5, which is for 36 stations over 100 yr and the 10 k EBCM run is used for the EOFs) several important indicators as a function of truncation level. Shown in the top row [Figs. 5(1a), 5(1b)] is the cumulative projection squared of a particular signal onto the retained modes. This is essentially the fraction of the variance of the signal explained by the number of modes retained.

$$\text{Fraction explained} = \frac{\sum_n^{n_{\text{trunc}}} S_n^2}{\sum_n^{n_{\text{total}}} S_n^2}. \quad (13)$$

This is a monotonic function approaching unity for all signals as  $n_{\text{trunc}}$  becomes large. One of our truncation criteria is that a good portion of the signal be captured in the retained modes. For example, in Fig. 5(1a) there is an abrupt increase at around retained mode 50; it would clearly be a mistake to truncate below this level—a wiser choice seems to be around 500.

Figs. 5(1b), 5(2b), etc., show the eigenvalues of the corresponding space–time EOF modes; these are the same for all columns of panels. The c panels show the individual contributions to  $\gamma^2$  (defined as the signal-to-noise ratio squared for the single variable regression). This quantity

TABLE 6. Same as Table 1 except for HadCM2 greenhouse gas (*G*) and greenhouse gas-plus-aerosols (*GA*) signal detection. Rows 1–2 are based on 72 global boxes (36 boxes have 100-yr data and 36 boxes have 50-yr data); rows 3–4 are based on 50-yr data (1944–93) for 72 detection boxes.

Signal	GFDLc		GFDLml		EBCM		ECHAM1/LSG		HadCM2	
	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$
<i>G</i>	5.28	0.59	5.87	0.46	4.87	0.39	4.83	0.42	4.13	0.35
<i>GA</i>	3.44	0.10	3.90	0.15	2.98	0.26	3.38	0.10	2.87	0.19
<i>G</i>	4.94	0.30	4.87	0.38	4.64	0.25	5.00	0.34	3.90	0.35
<i>GA</i>	3.25	0.35	2.71	0.49	2.80	0.53	2.58	0.41	2.60	0.44

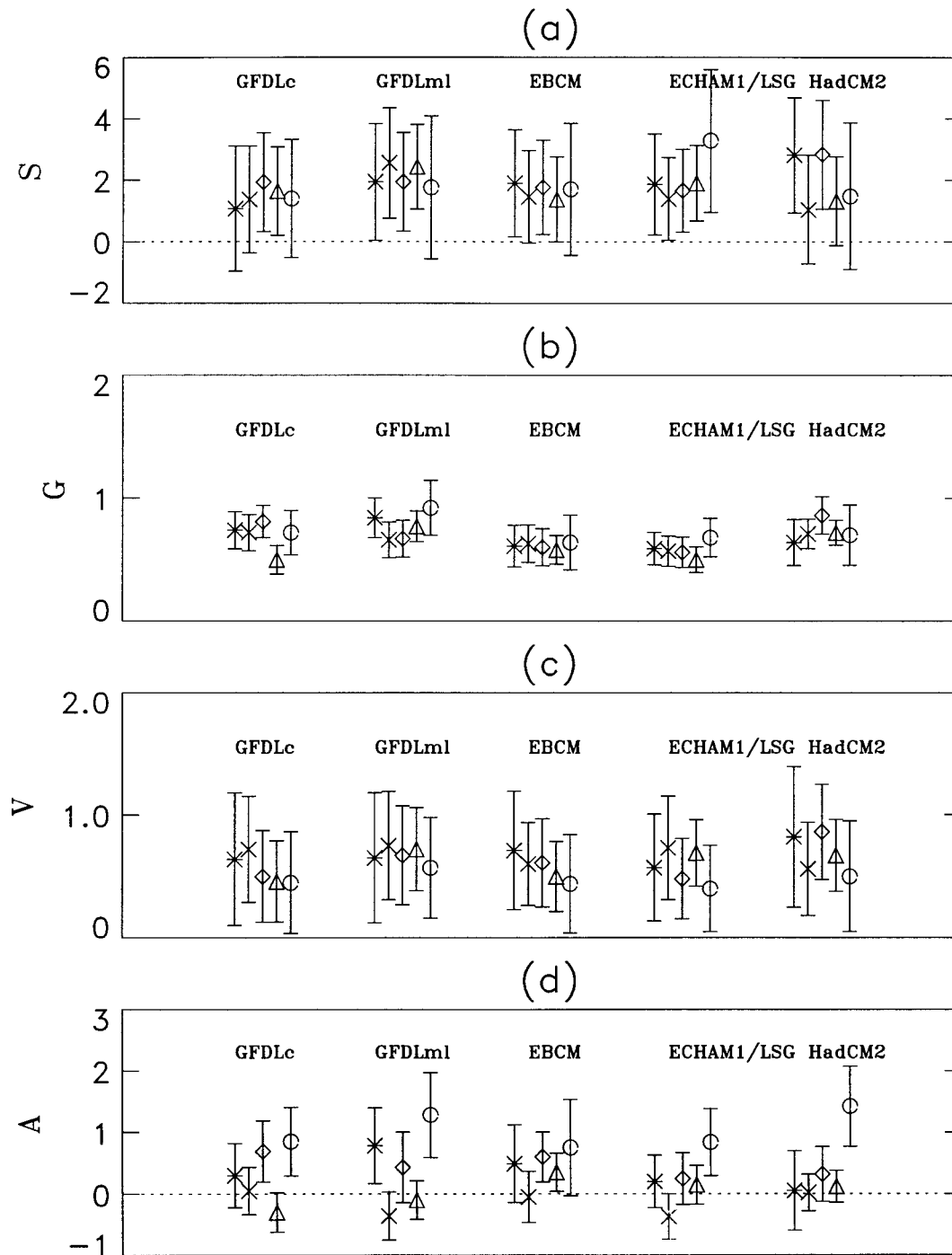


FIG. 3. Estimates of the signal amplitudes with 90% confidence interval indicated. These are grouped in clusters of five, each indicating a different record/site configuration. The groups are labeled by the control run used in the optimal weighting. (a) Is for the solar signal ( $S$ ); (b) indicates the greenhouse gas signal ( $G$ ); (c) is for the volcanic signal ( $V$ ); and (d) is for the anthropogenic aerosols ( $A$ ). The \* indicates that the estimate is based on 20 tropical stations with 100-yr records;  $\times$  is based on the 36 global stations as in NS98;  $\diamond$  is based on 43 tropical stations—20 with 100 yr and 23 with 50 yr of data;  $\blacktriangle$  is based on 72 global stations—36 with 100 yr and 36 with 50 yr of data;  $\circ$  is based on 72 stations with 50 yr of data (1944–93).

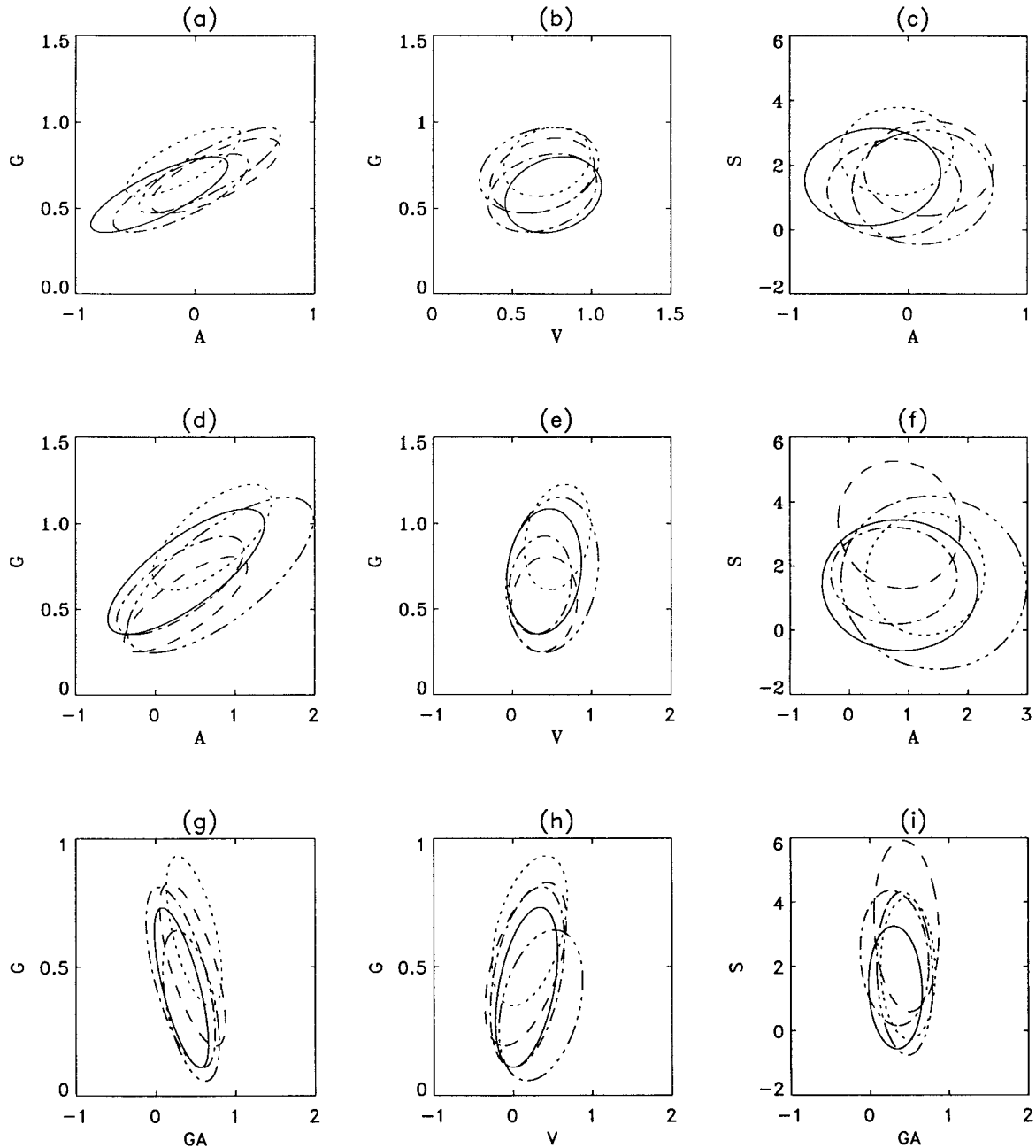


FIG. 4. Error ellipses of pairs of signals given five different model prescriptions for the natural variability: GFDLc (solid line); GFDLml (dotted line); MPI (dashed line); EBCM (dashed-dotted line); HadCM2 (dashed-dotted-dotted line). Here (a), (b), and (c) are EBCM signals for 72 global stations, 36 with 100 yr, 36 with 50 yr of data; (d), (e), and (f) are EBCM signals for 72 global sites all with 50 yr of data; (g), (h), and (i) are HadCM2  $G$  and  $GA$  and EBCM  $V$  and  $S$  signals for 72 global sites all with 50 yr of data.

does depend on the individual signal being considered. Note, for example, in Fig. 5(1c) the dramatic contributions from certain individual modes. These modes are related to temporal frequencies around  $11^{-1}$  cycles  $\text{yr}^{-1}$ .

Figures 5(1d), 5(2d), 5(3d), and 5(4d) show the cumulative signal-to-noise ratio squared as a function of trun-

cation level. In the case of the solar signal  $S$  we see a steady increase of cumulative  $\gamma^2$ , but its value is still below or near unity at level 1200. This is an indication that the sampling error is still substantial at all mode levels.

Figures 5(1e), 5(2e), 5(3e), and 5(4e) show the estimate of  $\alpha_S$  for each of the signals as a function of the

number of retained modes. It is interesting to examine Fig. 5(1e) for the  $S$  case. The sampling error is evident in the irregularity of the estimator. At truncation level 500, it is near unity.

By way of contrast, consider the estimate of  $\alpha_C$  in Fig. 5 column 2. Even at truncation level 50 we have captured most of the signal, but looking at Fig. 5(2c) we see that large contributions to  $\gamma^2$  are coming from modes near 400. We might agree on a truncation level of 500 also noting how stable the estimate of  $\alpha_C$  is in Fig. 5(2e). Its value is around 0.6.

A similar analysis holds for  $V$  in column 3. Truncation at around 500 leads to a rather stable estimate of about 0.7 for  $\alpha_V$ .

The bottommost panels (Figs. 5(1f), 5(2f), etc.) are a measure of the residual errors squared compared to the natural variability in the model control runs. This graph is independent of the signal being considered so the graphs are the same from one column to another. This is essentially the same test as applied in AT99 [Eq. (19)]. We plot the inverse of the  $\chi^2$  quantity

$$\hat{\chi}_{n_{\text{trunc}}}^2 = \frac{1}{n_{\text{trunc}}} \sum_{i=1}^{n_{\text{trunc}}} \frac{(y_i - \hat{y}_i)^2}{\lambda_i}, \quad (14)$$

where  $y_i - \hat{y}_i$  are the residuals of the regression process for mode  $i$  and  $\lambda_i$  is the corresponding eigenvalue from the control run. For very long control runs the quantity  $\hat{\chi}_{n_{\text{trunc}}}^2$  tends to a  $\chi_{n_{\text{trunc}}}^2$  distributed variate. This is nearly so for our 10 k run. If the eigenvalues are extracted from shorter runs (such as 1 k) there will not be so many degrees of freedom because there will be some correlation between the terms. The dashed lines in the figure indicate the 90% confidence band for this quantity. When the value is above the dashed line, it indicates that the model's natural variability is more than the natural variability of the residuals in the data. One may also tend to reject the null hypothesis that the model's natural variability is drawn from a population indistinguishable from that in nature. One can see that in our graphs Figs. 5(1f)–5(4f) the actual line always falls above the confidence limits at truncations levels below about 900, approaching it very slowly. This means that our confidence limits based upon this model for natural variability will be very conservative (error ellipses perhaps too large). Note also that if the reduced number of degrees of freedom due to sampling error are taken into account, the width of the "acceptable" region enlarges, but note that the dependence on the number of degrees of freedom is very small for large truncation level. The reader is to recall that our eigenvalues are sample estimates of the eigenvalues and these are likely to be high biased even in the 10 k yr run that was used in the construction of Fig. 5.

In arriving at numerical estimates (as shown in the tables) of the  $\alpha_s$  values we take three criteria for the truncation of the EOF sequence into account.

1) We do not wish to retain EOFs that contribute too

little to the variance of the natural variability. Small eigenvalues in the denominators can cause spurious errors and misleading indicators of signal to noise. The sum of retained EOF contributions to the variance is never allowed to exceed 99%. In other words, EOFs at the tail of the series that total one percent of the variance are never included. In the case of the 10 k run (Fig. 5) the contributions to  $\gamma^2$  (Figs. 5(1c), 5(2c), etc.) taper off as the truncation level becomes large. But when a 1 k run is used (not shown) we see rather large contributions to  $\gamma^2$  as the truncation level goes above about 500, indicating that the optimal truncation level is significantly affected by the length of the control run used. This phenomenon was also discussed by NS98.

- 2) The fraction of variance in explaining the signal squared summed over the domain should be large ( $>0.60$ ). This cumulative sum as function of truncation level is shown in Figs. 5(1a)–5(4a).
- 3) As a guide we also used the inverse  $\chi_{n_{\text{trunc}}}^2$  graph. We felt it was acceptable to truncate the EOF series when the inverse  $\chi^2$  lay above the acceptance region since this would lead to (larger) more conservative error ellipses. On the other hand, the level of natural variability in the stochastic EBCM is controlled entirely by the level of forcing noise in that model. Stevens tuned his model to have the correct level of variance at the seasonal cycle timescale. One could tune the model at longer timescales and thereby scale every eigenvalue by the same factor. This would bring the control run  $\chi^{-2}$  into agreement in Figs. 5(1f)–5(4f) and make the EBCM error ellipses smaller. Note that lowering the curve in these figures by a factor of say  $1/1.5$  shrinks the principle axes of the error ellipses by  $1/\sqrt{1.5}$ , causing the dash-dot ellipses of Figs. 5(4d) and 5(4f) to lie completely to the right of the  $A = 0$  vertical line.

In the tables and charts we used a different truncation level for each case based on the criteria above. But in all cases the resulting estimate of  $\hat{\alpha}_s$  was insensitive to truncation level over a wide range.

## 5. Conclusions

We have found strong evidence for the presence of the greenhouse and volcanic signals in the surface temperature data stream. In all the site/record-length configurations (25) we ran, we were able to reject the null hypothesis that the signal was absent at the 10% level. Somewhat surprising to us was the fact that these signals were weaker than our trial signal amplitudes, usually the most probable values were of the order of 0.5–0.7. For  $G$  the value 1.0 was included in the 90% band in only a few of the 25 cases.

The contribution due to the cooling effect of anthropogenic aerosols is rather more ambiguous. While the value 1.0 lies within the 90% confidence band for most

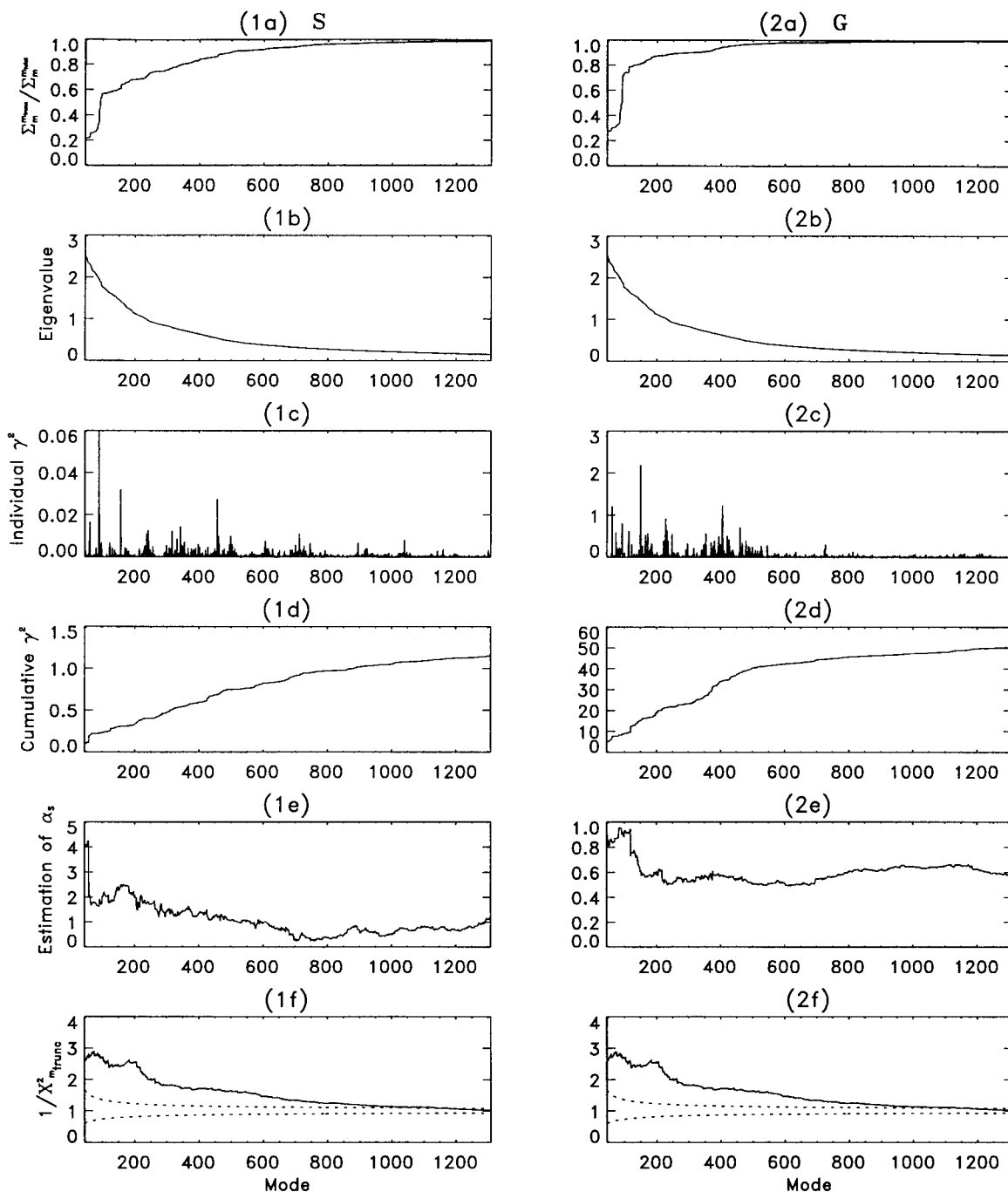


FIG. 5. (1a)–(1e) For solar cycle (S), (2a)–(2e) for greenhouse gas (G), (3a)–(3e) for volcanic (V), and (4a)–(4e) for aerosol (A) in the 36 global station case over and 100 yr. The space–time EOF modes are arranged in order of descending variance (EOFs from 10 k EBCM control run). (1a)–(4a) The normalized cumulative fraction of variance of the signal,  $\sum_n^{nunc} S_n^2$  with  $\sum_n^{all} S_n^2 = 1$ ; (1b)–(4b) indicate the eigenvalue of each spatial–temporal mode; (1c)–(4c) indicate the contributions to the SNR<sup>2</sup> =  $\gamma_n^2 = S_n^2/\lambda_n$  from the individual EOF modes; (1d)–(4d) indicate the cumulative  $\gamma^2 = \sum_n^{nunc} \gamma_n^2$ . (1e)–(4e) The cumulative estimate of  $\alpha$  including EOFs up to  $n_{nunc}$ . (1f)–(4f) The inverse of the  $\chi^2$  quantity in (14).



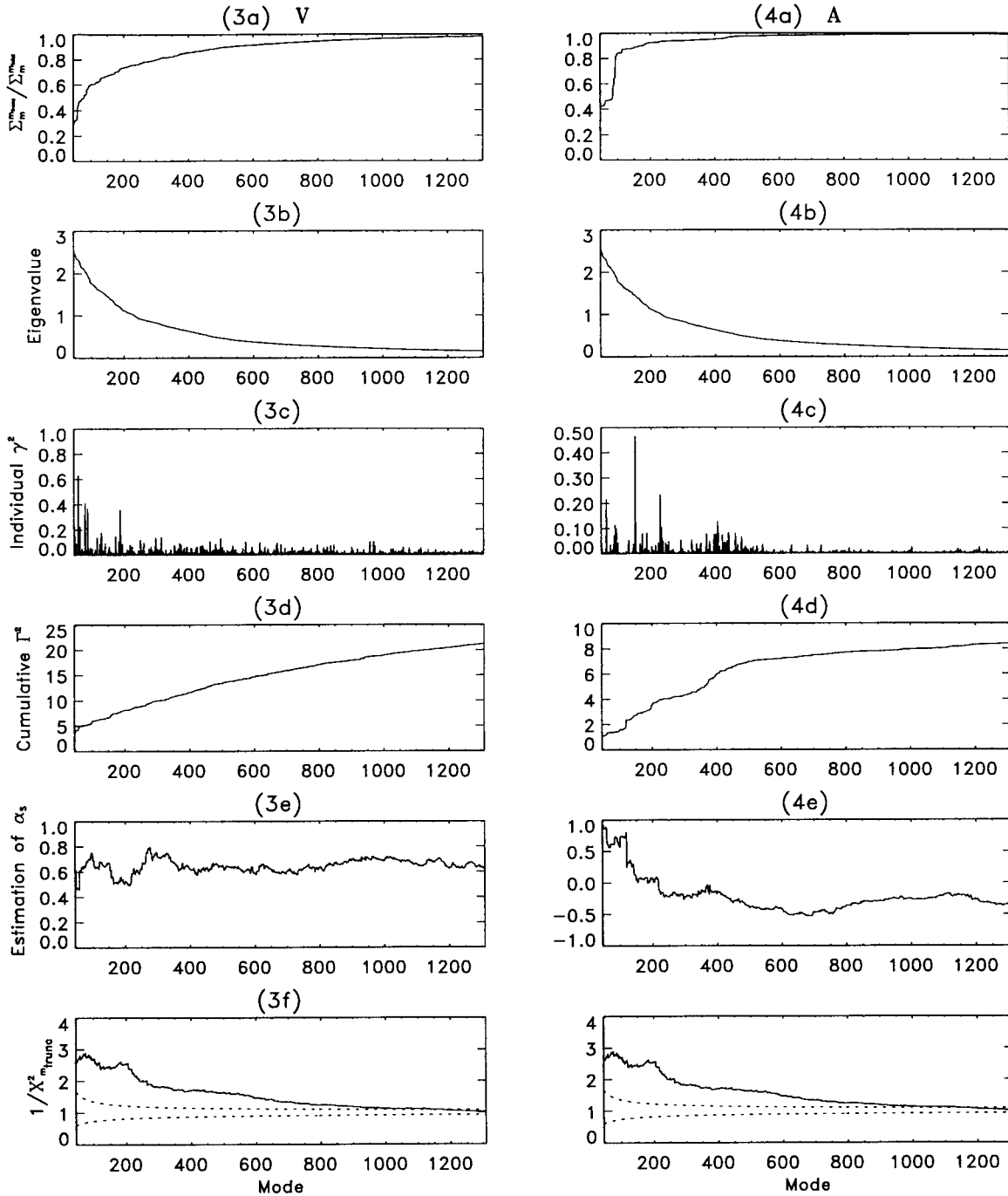


FIG. 5 (Continued)

of the runs in Fig. 3, so does the value 0.0 in about as many cases. A significant signal is found most prominently in the configurations using only the last 50 yr (1943–93); to be compatible with the Hadley Group we conducted some runs on the 50-yr period ending in 1996 and found no differences. The significant aerosol result was found to hold in 3 out of the 5 cases (choices of

natural variability model) we ran with the last 50 yr of data (circles in Fig. 3). This finding agrees with the most recent results available to us from the Hadley Group (P. Stott 2000, personal communication). We note that only a minor change in the noise forcing of the stochastic EBCM brings that natural variability in line to make it 4 out of 5 indicating significant aerosols in the records

of the last 50 yr. We believe the question of aerosol cooling and its amplitude in the data stream is still unsolved. Besides the uncertainties presented here due mainly to statistical sampling issues, there exist considerable difficulties in properly developing an aerosol signal that is robust across GCMs (see Hegerl et al. 2000). Even if we had a better idea of the direct effect of aerosols there remain the significant uncertainties of indirect effects that have to do with cloudiness changes induced by the presence of anthropogenic aerosols.

We venture to speculate on two possible causes of the weak aerosol signal as revealed in our analysis. One is the possible warming effects of some aerosols due to highly absorbing particles inside the droplets (e.g., Ackerman et al. 2000). Such warming effects have the same geographical pattern as the cooling and could to some extent lead to cancellation. Another potential offset is tropospheric ozone as suggested by Haywood et al. (1998). This greenhouse gas has a warming potential approximately the same strength as the cooling potential of the sulfate aerosol. Furthermore, the lifetime and source distribution of this gas is somewhat similar to that of anthropogenic aerosols. The similarity of the two would make it extremely difficult in a detection analysis to distinguish between them.

The solar signal is significantly above zero for all but 7 of the 25 configurations we ran (Fig. 3). The amplitude seems to lie in a 90% band between 0.0 and 4.0 with a most probable value of the order of 2.0. We believe this is an extremely interesting result, especially considering our crude representation of the solar signal. It is probably worthwhile to continue to explore opportunities to study this signal in various data streams and to improve our model-generated signals including some vertical dependences. Such an undertaking is, of course, difficult because of the many interactions of the solar beam with constituents of the upper atmosphere and the resulting alterations to the radiation fields.

Our finding that the greenhouse signal is somewhat weaker than expected might have serious implications. The finding is not because of the EBCM signal waveform because we have shown that the same result is obtained (in our analysis) when we use smoothed signals from the four-member HadCM2 ensemble average signals. Nor is the result dependent upon our style of EOF analysis. There are at least three potential physical explanations for the weak greenhouse signal: 1) the surface temperature field is less sensitive to greenhouse gas increases than we thought; 2) climate is sensitive, but the oceans are playing a greater or perhaps less obvious role in delaying the warming; 3) our signal waveforms are slightly incorrect in their space–time shape, leading to a low bias (smaller dot product of hypothesized and real patterns) in our estimates of strengths.

We believe that further improvements can be made in this type of approach to the problem. Such improvements might include better use of vertically dependent information and the inclusion of cyclostationary effects

such as the seasonal and diurnal cycles. Performance will obviously be enhanced with improved models and more data from observations.

*Acknowledgments.* The authors wish to thank Mark Stevens, Kwang Yul Kim, and Steven Leroy for many helpful discussions. Conversations with S. Tett, M. Allen, J. Mitchell, and G. Hegerl have also been helpful and stimulating. We especially thank the editor, F. Zwiers, and the anonymous referees for comments that led to significant improvements in the paper. The work was supported in part by grants from the NOAA Climate Detection Program and the DOE CHAMMP program. We are especially grateful to investigators at GFDL, MPI, and Hadley Centre for making public their simulation results.

## APPENDIX A

### Multiple Regression

From (1) we can rewrite

$$T_n^{\text{data}} = \sum_{s=1}^p \alpha_s S_{sn} + N_n, \quad (\text{A1})$$

where  $\alpha_1, \dots, \alpha_p$  are the signal strengths,  $s = 1, \dots, p$ ;  $S_{sn}$  are the projections of signal  $s$  onto the EOF mode  $n$ ; and  $N_n$  are the climate noise projections onto EOF mode  $n$ , assumed to have pdf  $\sim N(0, \lambda_n)$ .

The likelihood function can be written

$$L(\alpha_1, \dots, \alpha_p) = \prod_n \frac{1}{\sqrt{2\pi\lambda_n}} \exp\left\{-\frac{1}{\lambda_n} \left(T_n^{\text{data}} - \sum_s \alpha_s S_{sn}\right)^2\right\}.$$

We find the maximum likelihood estimators (MLE) by maximizing  $L(\bar{\alpha})$  with respect to its variables  $\alpha_1, \dots, \alpha_p$ . Let  $\mathbf{W}$  be a diagonal matrix containing the weights (inverses of the  $\lambda_n$ )

$$\mathbf{W}_{nn'} \equiv \frac{\delta_{nn'}}{\lambda_n}. \quad (\text{A2})$$

The normal equations for the maximum likelihood estimation of the amplitudes are

$$\sum_{s'} (S_s \cdot \mathbf{W} \cdot S_{s'}) \hat{\alpha}_{s'} = S_s \cdot \mathbf{W} \cdot T^{\text{data}}. \quad (\text{A3})$$

It is noteworthy that the maximum likelihood method places the eigenvalues under each term of the sum of squared errors. This is equivalent to the “prewhitening” transformation described by AT99 and in the text by Mardia et al. (1979). It is also the origin of the ubiquitous metric  $\mathbf{W}$  that appears between all “dot” products. The prewhitening transformation is a little more satisfactory, but the MLE approach may be less mysterious to some readers.

We can define the matrix

$$\mathbf{\Gamma}_{ss'} \equiv S_s \cdot \mathbf{W} \cdot S_{s'} = \sum_n \frac{S_{sn} S_{s'n}}{\lambda_n} \quad (\text{A4})$$

and after solving for  $\alpha_s$  in the normal equations above, we can show that

$$\text{cov}(\alpha_s, \alpha_{s'}) = (\mathbf{\Gamma}^{-1})_{ss'}. \quad (\text{A5})$$

Note that the elements of  $\mathbf{\Gamma}$  are the multidimensional generalization of  $\text{SNR}^2$  introduced in our earlier publications (North et al. 1995; NS98).

## APPENDIX B

### Filter Interpretation versus Regression

In NS98 as here there were four signals. The solution given by NS98 was to sum the signal vectors of the three signals not of interest and find the component of the data stream perpendicular to that vector sum. Unfortunately, this procedure depends (perhaps sensitively) on the amplitudes of the other three signals, which are of course unknown. There is a way around this problem. Since the dimension of the EOF space is at least four, it is possible to find a projection operator that annihilates all vectors parallel to the other three signal vectors simultaneously. We proceed now to find such a projection operator (filter) that will cleanse the data stream of components parallel to the signals not of interest regardless of their amplitudes. We can then find an estimate of the signal amplitude of interest without fear of leakage of those unwanted signals into the filtered data stream. Naturally, the amplitude of the signal of interest remaining in the filtered data stream may be considerably diminished because some of its components will be parallel to those directions being eliminated. Because of this reduced amplitude surviving the prefilter, we can expect a smaller SNR than might have obtained otherwise. It is pleasing to find in the end that the formula for the signal estimates based on this scheme are precisely the amplitudes that we would have obtained from the standard multiple regression algorithm. A short proof of this is sketched below.

The key to this idea lies in the construction of the dual basis for the unit vectors that lie along the directions of the signal vectors. First, it is good to partition the space of EOFs into groups of four (for the four signal case). Take a particular one of these partition foursomes (later the estimation results from each foursome can be optimally combined). There are four unit vectors that lie along the signals:  $\mathbf{a}_s$ ,  $s = 1, \dots, 4$ , which are not necessarily mutually orthogonal, but no two are parallel either—it is a skewed basis set. We can find (using, e.g., the Gram–Schmidt process) another set of (not unit) vectors  $\tilde{\mathbf{a}}_{s'}$ ,  $s' = 1, \dots, 4$ , called the dual basis set such that  $\tilde{\mathbf{a}}_{s'} \cdot \mathbf{W} \cdot \mathbf{a}_s = 0$ ,  $s \neq s'$ , and  $\tilde{\mathbf{a}}_s \cdot \mathbf{W} \cdot \mathbf{a}_s = 1$ ,  $s = 1, \dots, 4$ ; where we have introduced the metric  $\mathbf{W}$  for reasons that will be clear momentarily.

Taking the signal vectors to be unit vectors, the data vector in this four-dimensional subspace can be written:

$$\mathbf{D} = \sum_s \alpha_s \mathbf{a}_s + \mathbf{N}, \quad (\text{B1})$$

where again the bold face indicates that the EOF indices have been suppressed. By taking the dot product of  $\mathbf{D}$  with  $\tilde{\mathbf{a}}_s$ , (with respect to the metric  $\mathbf{W}$ ) we project the parts of the data stream that are simultaneously perpendicular to the other signals ( $s \neq s'$ ). In fact, this projection is an unbiased estimator of  $\alpha_s$ :

$$\hat{\alpha}_s = \tilde{\mathbf{a}}_s \cdot \mathbf{W} \cdot \mathbf{D} \quad (\text{B2})$$

as can be shown by taking the expectation of it and using  $\tilde{\mathbf{a}}_s \cdot \mathbf{a}_{s'} = \delta_{s'}$ , noting that  $\langle \mathbf{N} \rangle = 0$ .

The metric factor  $\mathbf{W}$  has the property of assuring that the expectation of the “square” of the residuals  $\langle (\mathbf{D} - \sum_s \alpha_s \mathbf{a}_s) \cdot \mathbf{W} \cdot (\mathbf{D} - \sum_s \alpha_s \mathbf{a}_s) \rangle = 1$ , since  $\langle \mathbf{N} \cdot \mathbf{W} \cdot \mathbf{N} \rangle = 1$ .

It is convenient to express our estimator in terms of the  $\mathbf{a}_s$  rather than the  $\tilde{\mathbf{a}}_{s'}$ . The  $\mathbf{a}_s$  and the  $\tilde{\mathbf{a}}_{s'}$  are related by

$$\tilde{\mathbf{a}}_s = \sum_{s'} (\mathbf{F}^{-1})_{ss'} \mathbf{a}_{s'}, \quad (\text{B3})$$

where  $\mathbf{F}_{ss'} = \mathbf{a}_s \cdot \mathbf{W} \cdot \mathbf{a}_{s'}$ . The unbiased estimator becomes

$$\hat{\alpha}_s = \tilde{\mathbf{a}}_s \cdot \mathbf{W} \cdot \mathbf{D} = \sum_{s'} (\mathbf{F}^{-1})_{ss'} \mathbf{a}_{s'} \cdot \mathbf{D}. \quad (\text{B4})$$

We can identify the terms in the multiple regression estimator of  $\alpha_s$ :

$$\mathbf{S} \cdot \mathbf{W} \cdot \mathbf{S} \rightarrow \mathbf{F} \quad (\text{B5})$$

$$\mathbf{S}_{s'} \cdot \mathbf{W} \cdot \mathbf{T}^{\text{data}} \rightarrow \mathbf{a}_{s'} \cdot \mathbf{W} \cdot \mathbf{D}. \quad (\text{B6})$$

The above alternative view of the regression formula (we did not show that it was the least mean squares estimator that is the starting point of the conventional derivation, but this is trivially true) shows that the filter interpretation is useful and, in fact, valid. In estimating the strength of signal  $s$  we filter away all the components of the data stream that are perpendicular to the other signals,  $s' \neq s$ , simultaneously by use of the dual basis set. Having established this geometrical interpretation, we can see that the matrix  $\mathbf{\Gamma}_{ss}^2$  is the  $\text{SNR}^2$  for the component of the signal that is mutually perpendicular to the other signals simultaneously.

The steps in the above procedure solve the problem for an EOF subset of four. We now simply combine all the subsets of four in an optimal combination, weighted inversely by their variances. It takes a few steps to prove that this procedure leads exactly to the regression formula for optimal estimation of  $\alpha_s$  for any multiple of four EOFs. By way of analogy, consider the estimation of the temperature of a bath with  $N$  thermometers (discussed in NS98). We can combine all of the readings weighted inversely by their variances, or we can take optimally weighted subsets of them, weighing each subset inversely by the variance of that subset. The final result is the same no matter how the subsets are partitioned and combined if they are always combined optimally.

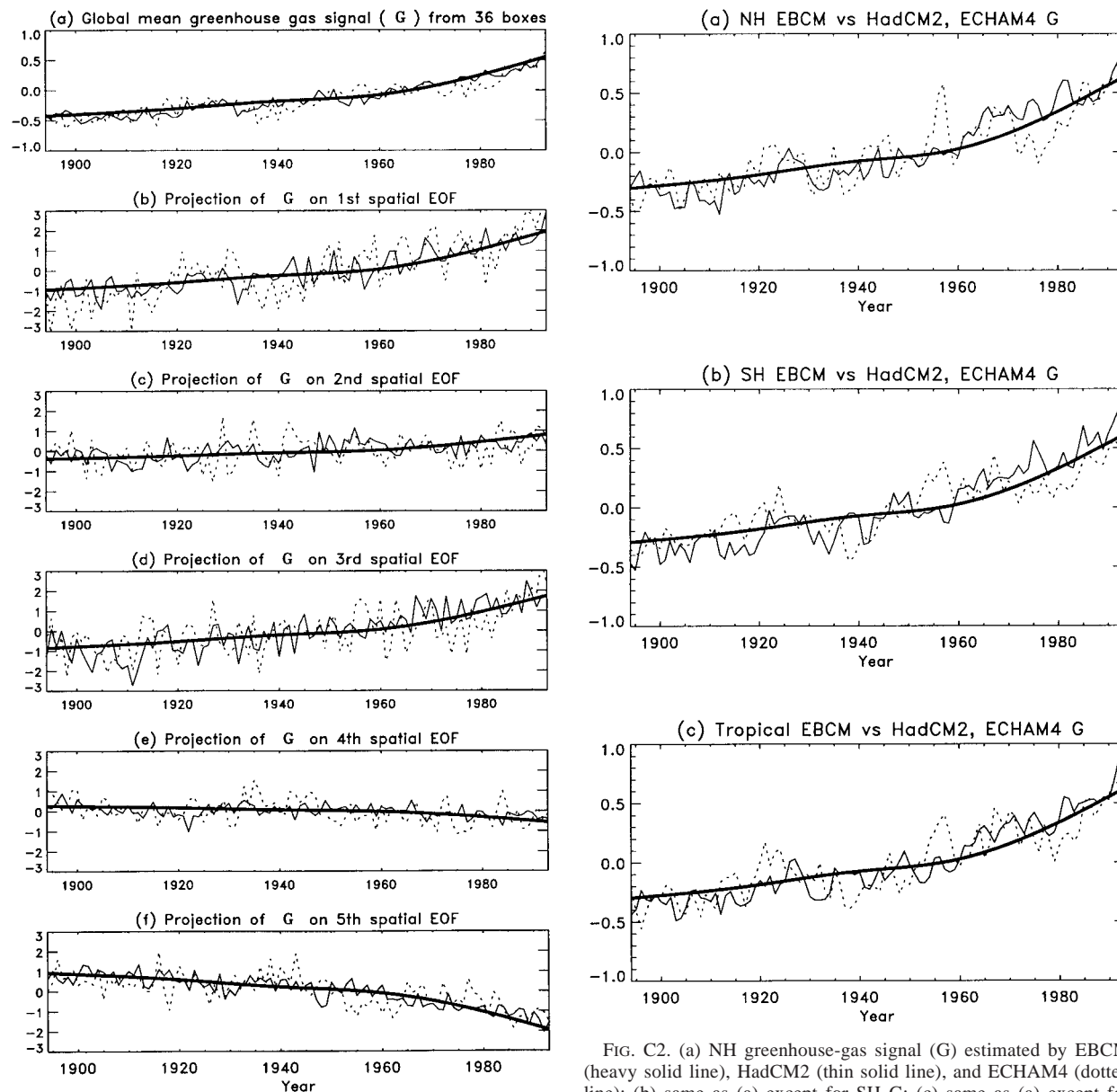


FIG. C1. (a) Global greenhouse-gas signal ( $G$ ) estimated from 36 detection boxes by EBCM (heavy solid line), HadCM2 (thin solid line), and ECHAM4 (dotted line); (b) projection of 36 box signal on first spatial EOF estimated by 10 K EBCM control run; (c) projection of 36 box signal on second spatial EOF estimated by 10 k EBCM control run.

## APPENDIX C

### GCM versus EBCM Signals

Comparison of the EBCM greenhouse signal with that of the HadCM2 were presented in Fig. 2, but these results might be difficult to interpret. We have aggregated the signals in some different ways to facilitate the comparison. Figure A1a shows the 36 box (global) av-

FIG. C2. (a) NH greenhouse-gas signal ( $G$ ) estimated by EBCM (heavy solid line), HadCM2 (thin solid line), and ECHAM4 (dotted line); (b) same as (a) except for SH  $G$ ; (c) same as (a) except for tropical  $G$ .

erage versus time for  $G$  as computed by the EBCM (heavy solid line) and the same for the HadCM2 (thin solid line) and ECHAM4 (dotted line). Figures C1b–C1f indicate the time dependences of the projections of these signals onto the first five spatial EOFs (computed from the 10 k run of our stochastic EBCM) of the same models (note that the sign of the signal here is arbitrary, since the sign of an EOF is not determined). Figure C2a shows a NH average taken from the 36 boxes for  $G$  from the three models. Similarly Figs. C2b,c show the SH and tropical averages. Similar agreement obtains for the GA case (not shown). Figure C3a shows an index of the  $G$  signal computed from each of the EBCM

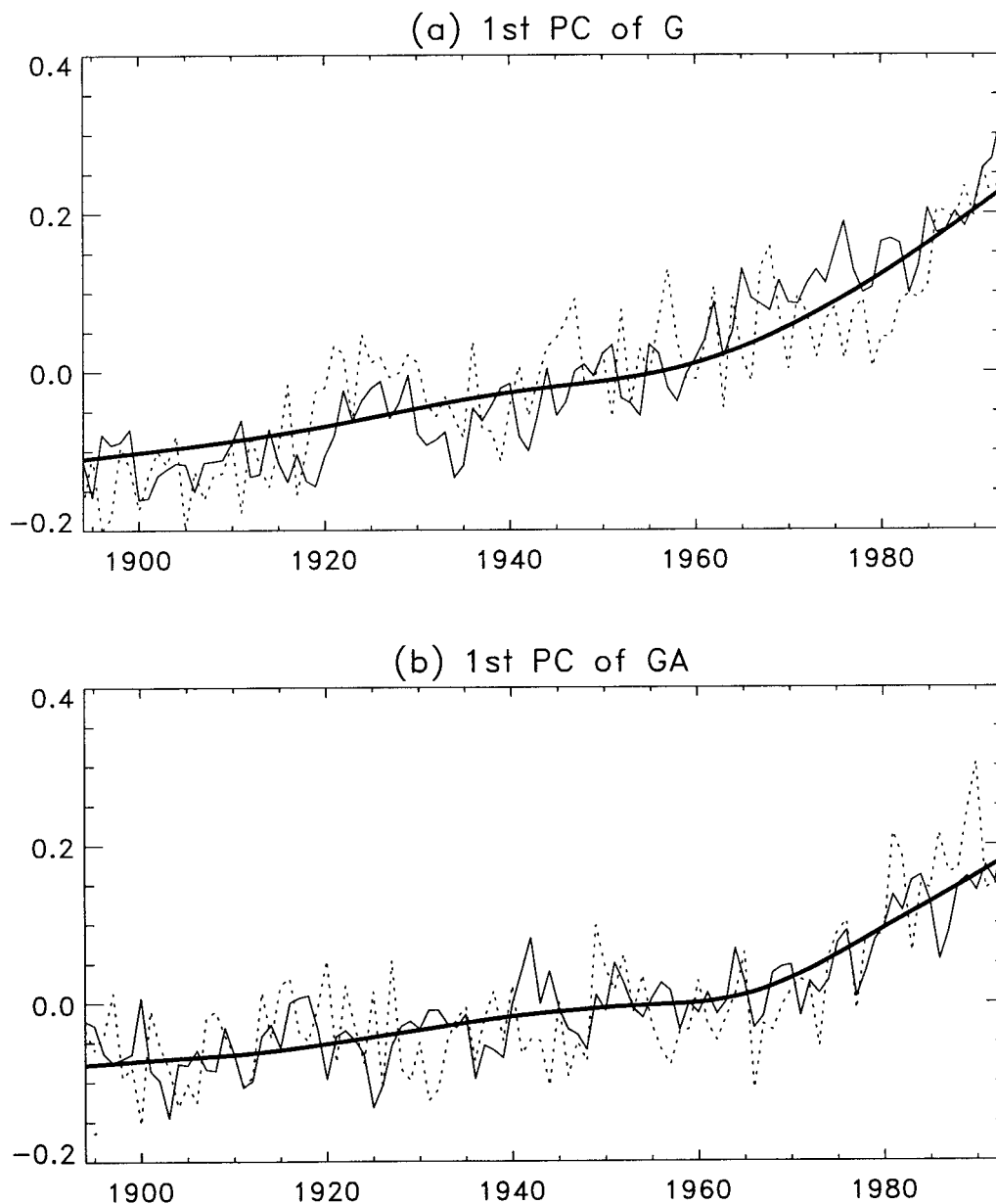


FIG. C3. (a) First principle component time series of annual mean climate change signal for greenhouse-gas-only, forcing from EBCM (heavy solid line), HadCM2 GCM (light solid line), ECHAM4 GCM (dotted line); (b) same as (a) except for greenhouse-gas-plus-aerosol forcing.

(heavy solid), HadCM2 (light solid), and ECHAM4 (dotted). Figure C3b shows the same for GA as defined in the text. It is apparent that the EBCM is an excellent approximation to the signals generated by the GCMs at least at these large spatial scales. Certainly at these scales it compares as favorably to either of them as they do to each other. We have conducted similar analyses with other GCMs (not shown), finding the same results. Based upon these comparisons we conclude that the EBCM signals are adequate for the tasks set forth in our paper.

## APPENDIX D

### Monte Carlo Check

In such a complicated regression procedure, it is useful as a check to pass a synthetic data stream through the filter to see if there are unforeseen biases and to check that the confidence ellipses have been accurately computed. This is possible since we have a 10 k yr run with our noise-forced EBCM. The confidence ellipse is as described in section 4b. The shape, size, and centroid of the ellipse depend on the data used in computing the

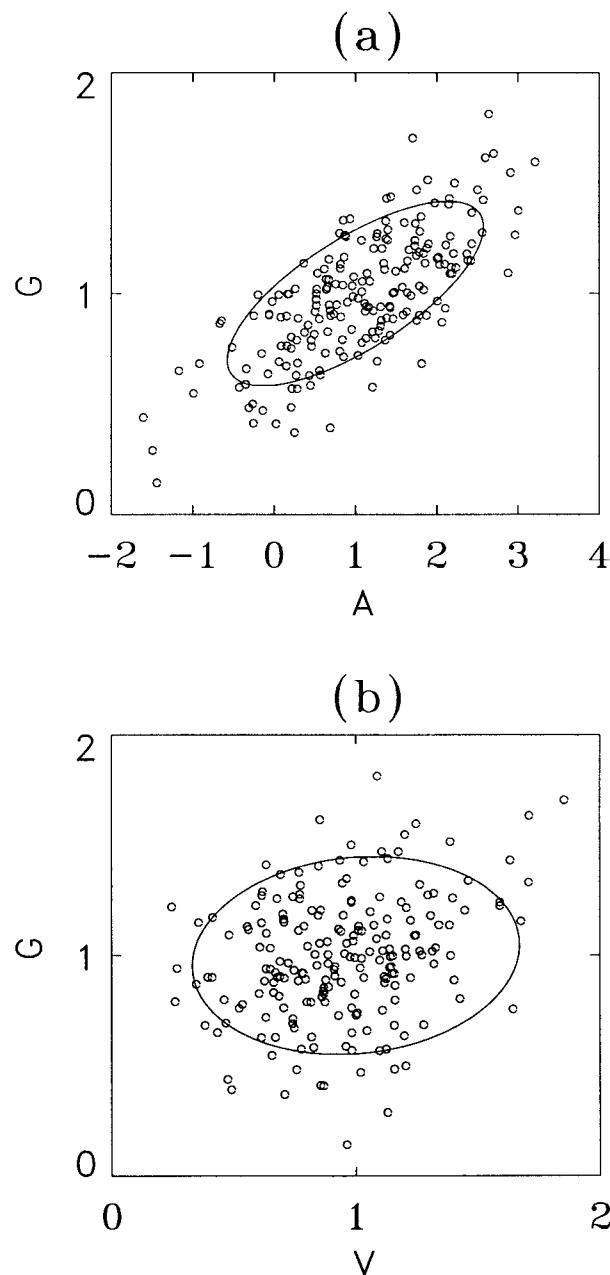


FIG. D1. (a) Scatterplot of Monte Carlo studies and 90% error ellipse of detection studies for pair of signal  $G$ - $A$  for 72 boxes all with 50-yr (1944-93) observational data. In Monte Carlo studies, the artificial data is constructed by adding 200 50-yr EBCM control run and four EBCM signals  $S$ ,  $G$ ,  $V$ , and  $A$ . The truncated eigenmode is 500 in the 10 k yr control run of EBCM. (b) Same as (a) except for pair of signal  $G$ - $V$ .

covariance matrix. We used the same covariance matrix as in Fig. 4, except that we translated its centroid to the point (1, 1) in the diagram; otherwise, the shape and size of the ellipse was the same.

We have passed artificial data streams with our natural variability and our four hypothesized signals through

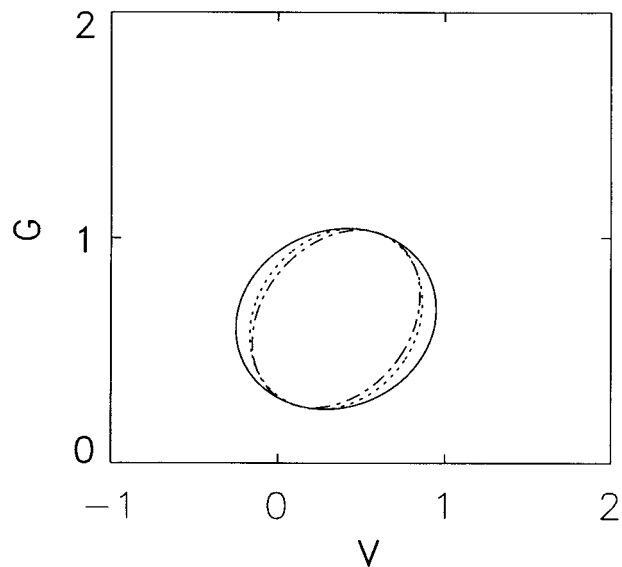


FIG. F1. The 90% error ellipse of detection studies for pair of signal  $G$ - $V$  for 72 boxes all with 50 yr (1944-93) observational data. The three error ellipses are obtained by three different EOF truncation levels: 300 (solid line), 500 (dashed line), and 700 (dashed-dotted line). In this study, the EOFs are from the 10 k yr control run of EBCM.

the filter. Figure D1 shows examples (50-yr interval, 72 boxes) of two 90% confidence ellipses with 200 realizations passed through the filter. One can see that of the order of 40 (~20%) of the points fall outside the ellipse in each panel. In this case the truncation level of the EOF expansion was 500. We show these examples mainly to illustrate that the estimates are not noticeably biased. But the ellipses are a bit small. We performed the same exercise with covariance matrices based upon artificial data. In each case the shape and size of the ellipse varied slightly. Over many such realizations the average number of points outside was indeed 10%.

## APPENDIX E

### Sensitivity to EOF Sampling Error

The adequacy of a 1000-yr control run for space-time EOF estimation on a 100-yr interval is open to question (see AT99). We performed a few tests to check our results against one run we have with the EBCM, which is 10 k yr in length. One interesting test is the sensitivity of the error ellipses to one 1000-yr segment being used for computing the eigenvalues and eigenvectors and another 1000-yr segment being used for the computation of the error ellipse. We did this many times with the 10 1000-yr segments available from the EBCM control run and found that the location of the centroid of error ellipses as well as their sizes and shapes were hardly affected. Certainly this contribution to the uncertainty is an order of magnitude less than the differences re-

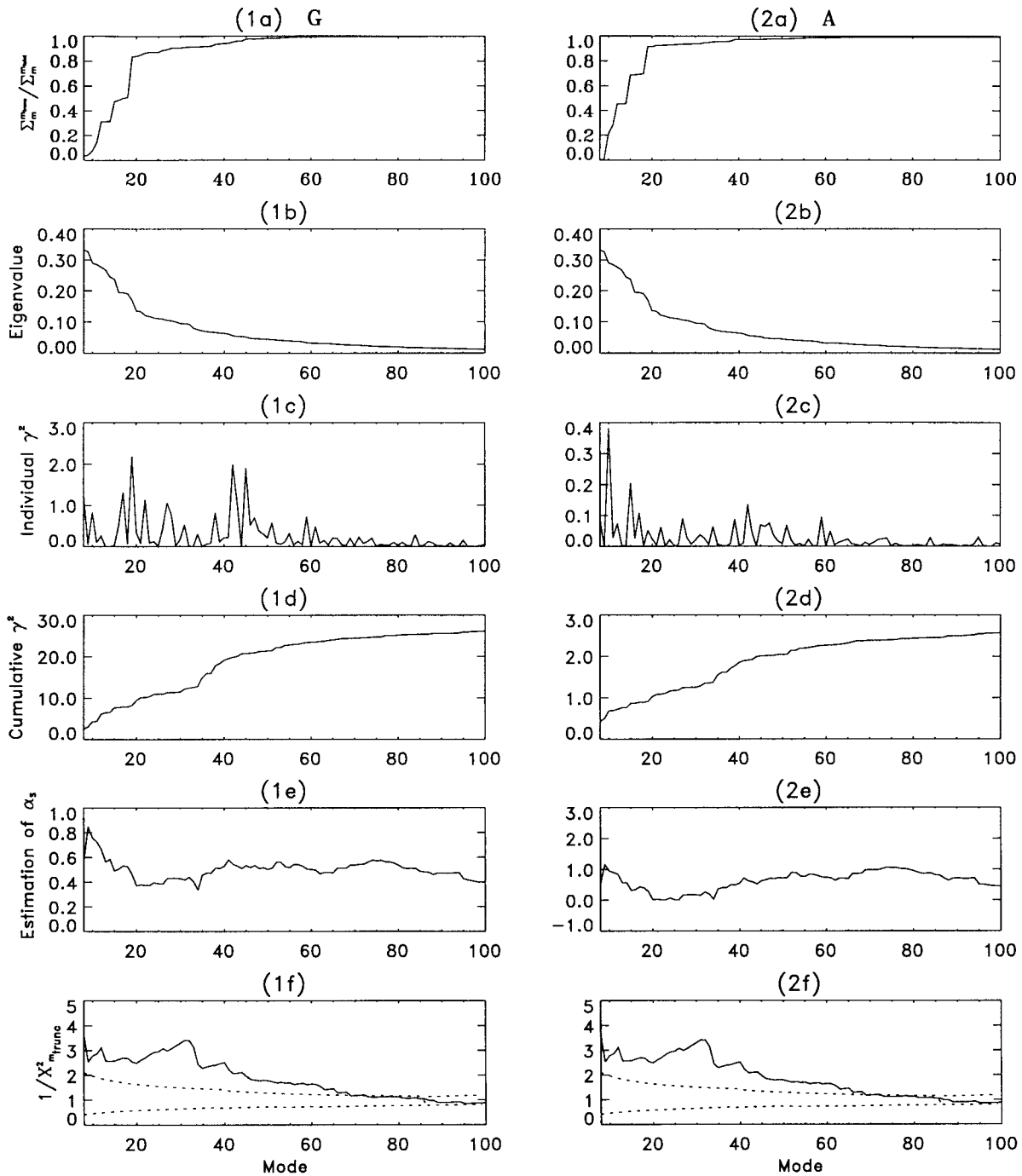


FIG. G1. Same as Fig. 5 in the main text except for the decadal smoothed data case. Only G and A are detected in the observed data. The climate signal, observed data, and natural variability are equally smoothed.

sulting from the use of eigenvalues and EOFs from one model to another.

One sensitivity that does arise when using only 1 k yr of data is the truncation level. In Fig. 5c one sees the contributions to  $\gamma^2$  from individual modes. In this illustration, which used 10 k yr of data to compute the

EOFs, the contribution from large ( $>600$ ) is very small. However, in most cases using only 1 k yr of control run data, the contributions to  $\gamma^2$  begin to increase for levels of truncation beyond 600. This is because the eigenvalues appearing in the denominators are very small in this range. The appearance of large contributions to  $\gamma^2$

for larger truncation levels is a signal to truncate the series. These findings also tell us that it is advisable to have longer control runs in order to allow larger truncation levels, thereby increasing the power of the estimate (however, see appendix F).

## APPENDIX F

### Sensitivity to Truncation Level

From the analysis panels (Fig. 5 in the text), one can easily see the fluctuation in the estimate of  $\alpha_{G,A,S,V}$  as one varies the truncation level. As the truncation level is increased the error ellipses will shrink because more estimators are being included. This will be true so long as there is some nonnegligible signal projected onto the additional modes being included. Figure F1 gives an indication of the sensitivity of the error ellipses for 200, 400, and 600 levels of truncation for the  $G$  and  $V$  signal amplitudes being estimated simultaneously. The ellipses are essentially indistinguishable for truncations beyond about 650. In these cases the EOFs were computed from the 10 k EBCM run.

It is important to reflect on the use of EOFs in this problem. The regression formula for the amplitudes is given by (6). In that formula the matrix  $\mathbf{W}$  is simply the inverse of the sample covariance matrix  $\mathbf{K}$ . By going to the EOF representation we simply are performing a rotation of the coordinate system from the space–time discrete points to the EOF basis set. This conveniently casts the matrix  $\mathbf{W}$  into diagonal form, making the estimate  $\hat{\alpha}$ , a series of (asymptotically, for large sample) monotonically decreasing terms. In this diagonal form we have a natural means of making choices about truncation (reducing the dimension of the space under consideration). Just as in ordinary estimation theory after the truncation choice is made, we must go back and renormalize the estimator so that it is theoretically unbiased. That is to say that each independent estimator is inversely weighted by its own error variance, but when estimators are combined, the linear combination is normalized so that the result is unbiased if each individual term is unbiased. To the extent that the sample EOFs are still slightly correlated or if the sample eigenvalues are biased (which they are almost certainly), the estimates of signal strengths will be suboptimal, but not biased.

## APPENDIX G

### Sensitivity to Smoothing

One issue that seems to trouble some readers is the huge number of space–time EOFs retained in our study. As discussed in the text we choose to retain these many EOFs to gain the advantage of data averaging at the end of the estimation process rather than at the beginning. One might choose to lump data into decadal averages or trends at the beginning, but this has the disadvantage

of giving up the ability to estimate the volcanic and solar contributions. Lumping data at the beginning has the effect of reducing the size of the EOF space with which one is dealing, but this must be compared to the loss of ability to detect the volcanic and solar signals. Of course, if one has only a few realizations from a GCM run decadal averaging has the additional benefit of smoothing the signal waveforms and thereby reducing the error committed. For comparison we have redone our analysis by lumping our signals and data into 11-yr moving averages.

Following the mode analysis as outlined in Fig. 5, we perform the same analysis on 11-yr moving averaged inputs (36 stations, 100 yr of data; EBCM signals, variability from the 10 k yr run) as shown in Fig. G1. In the truncation range 20–40 we find a value of  $\hat{\alpha}_G \approx 0.4$ . But if we include up to 70 modes the values increase to  $\approx 0.5$ . From the point of view of the  $\chi^2$  test we might prefer the 70 mode retention because this tends to pass the test according to Fig. G1e. Note that at this level the cumulative  $\gamma^2$  is of the order of 20.

As for the aerosols, at truncation level 70, we find an estimated amplitude of the order of unity with  $\gamma^2 \approx 2.4$ , a reasonably clear and robust signal.

## REFERENCES

- Ackerman, A. S., O. B. Toon, D. E. Stevens, A. J. Heymsfield, V. Ramanathan, and E. J. Welton, 2000: Reduction of tropical cloudiness by soot. *Science*, **288**, 1042–1047.
- Allen, M. R., and S. F. B. Tett, 1999: Checking for model consistency in optimal fingerprinting. *Climate Dyn.*, **15**, 419–434.
- Barnett, T. P., 1986: Detection of changes in the 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. Schlesinger, Ed., Elsevier, 559–568.
- , and M. E. Schlesinger, 1987: Detecting changes in global climate induced by greenhouse gases. *J. Geophys. Res.*, **92**, 14 772–14 780.
- , and Coauthors, 1999: Detection and attribution of recent climate change: A status report. *Bull. Amer. Meteor. Soc.*, **80**, 2631–2659.
- 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 to detect climate change. *J. Atmos. Sci.*, **43**, 1694–1710.
- Hasselmann, K., 1979: On the signal-to-noise problem in atmospheric response studies. *Meteorology of Tropical Oceans*, D. B. Shaw, Ed., Royal Meteorological Society, 251–159.
- , 1993: Optimal fingerprints for the detection of time-dependent climate change. *J. Climate*, **6**, 1957–1971.
- Haywood, J. M., M. D. Schwarzkopf, and V. Ramaswamy, 1998: Estimates of radiative forcing due to modeled increases in tropospheric ozone. *J. Geophys. Res.*, **103**, 16 999–17 007.
- Hegerl, G. C., H. von Storch, K. Hasselmann, B. D. Santer, U. Cubasch, and P. D. Jones, 1996: Detecting greenhouse gas-induced climate change with an optimal fingerprint method. *J. Climate*, **9**, 2281–2306.
- , P. A. Stott, M. R. Allen, J. F. B. Mitchell, S. F. B. Tett, and U. Cubasch, 2000: Optimal detection and attribution of climate



- change: Sensitivity of results to climate model differences. *Climate Dyn.*, **16**, 737–754.
- Johns, T. C., R. E. Carnell, J. F. Crossley, J. M. Gregory, J. F. B. Mitchell, C. A. Senior, S. F. B. Tett, and R. A. Wood, 1997: The second Hadley Centre coupled ocean–atmosphere GCM: Model description, spin-up, and validation. *Climate Dyn.*, **13**, 103–134.
- Jones, P., and K. R. Briffa, 1992: Global surface air temperature measurements during the twentieth century: Spatial temporal and seasonal details. *Holocene*, **2**, 165–179.
- Kim, K. Y., and Q. Wu, 2000: Optimal detection using cyclostationary EOFs. *J. Climate*, **13**, 938–950.
- Leroy, S., 1998: Detecting climate signals: Some Bayesian aspects. *J. Climate*, **11**, 640–651.
- Mardia, K. V., J. T. Kent, and J. M. Bibby, 1979: *Multivariate Analysis*. Academic Press, 521 pp.
- 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., 1984: Empirical orthogonal functions and normal modes. *J. Atmos. Sci.*, **41**, 879–887.
- , and K. Y. Kim, 1995: Detection of forced climate signals. Part II: Simulation results. *J. Climate*, **8**, 409–417.
- , and M. J. Stevens, 1998: Detecting climate signals in the surface temperature record. *J. Climate*, **11**, 563–577.
- , K. Y. Kim, S. S. P. Shen, and J. W. Hardin, 1995: Detection of forced climate signals. Part I: Filter theory. *J. Climate*, **8**, 401–408.
- Roeckner, E., L. Bengtsson, J. Feichter, J. Lelieveld, and H. Rodhe, 1999: Transient climate change simulations with a coupled atmosphere–ocean GCM including the tropospheric sulfur cycle. *J. Climate*, **12**, 3004–3032.
- Stevens, M. J., 1997: Optimal estimation of the surface temperature response to natural and anthropogenic climate forcings over the past century. Ph.D. dissertation, Texas A&M University, 157 pp.
- , and G. R. North, 1996: Detection of the climate response to the solar cycle. *J. Atmos. Sci.*, **53**, 2594–2608.
- Stott, P. A., S. F. B. Tett, G. S. Jones, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell, 2001: Attribution of twentieth century temperature change to natural and anthropogenic causes. *Climate Dyn.*, **17**, 1–21.
- Tett, S. F. B., J. F. B. Mitchell, D. E. Parker, and M. R. Allen, 1996: Human influence on the atmospheric vertical temperature structure: Detection and observations. *Science*, **247**, 1170–1173.
- , T. C. Johns, and J. F. B. Mitchell, 1997: Global and regional variability in a coupled AOGCM. *Climate Dyn.*, **13**, 303–323.
- , P. A. Stott, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell, 1999: Causes of twentieth-century temperature change near the earth's surface. *Nature*, **399**, 569–572.