

## Detecting Climate Signals in the Surface Temperature Record

GERALD R. NORTH AND MARK J. STEVENS

*Climate System Research Program, Texas A&M University, College Station, Texas*

(Manuscript received 21 October 1996, in final form 10 July 1997)

### ABSTRACT

Optimal signal detection theory has been applied in a search through 100 yr of surface temperature data for the climate response to four specific radiative forcings. The data used comes from 36 boxes on the earth and was restricted to the frequency band  $0.06\text{--}0.13$  cycles  $\text{yr}^{-1}$  (16.67–7.69 yr) in the analysis. Estimates were sought of the strengths of the climate response to solar variability, volcanic aerosols, greenhouse gases, and anthropogenic aerosols. The optimal filter was constructed with a signal waveform computed from a two-dimensional energy balance model (EBM). The optimal weights were computed from a 10000-yr control run of a noise-forced EBM and from 1000-yr control runs from coupled ocean–atmosphere models at Geophysical Fluid Dynamics Laboratory (GFDL) and Max-Planck Institute; the authors also used a 1000-yr run using the GFDL mixed layer model. Results are reasonably consistent across these four separate model formulations. It was found that the component of the volcanic response perpendicular to the other signals was very robust and highly significant. Similarly, the component of the greenhouse gas response perpendicular to the others was very robust and highly significant. When the sum of all four climate forcings was used, the climate response was more than three standard deviations above the noise level. These findings are considered to be powerful evidence of anthropogenically induced climate change.

### 1. Introduction

There are a number of different approaches presently being taken toward solving the problem of detecting forced climate change (Hasselmann 1979, 1993, 1997; Bell 1982, 1986; North et al. 1995; North and Kim 1995; Santer et al. 1995; Santer et al. 1996; Hegerl et al. 1996; Hegerl et al. 1997; Stevens and North 1996). The mathematical basis of these approaches share some commonality (Hegerl and North 1997), but the methods by which these techniques are applied and the data they are applied to differ widely among the various research groups. In the event that these different research groups conclude that evidence of climate change has been detected, or not detected, the credibility of their arguments will be enhanced by the use of different approaches to the problem. In fact, the acceptance of such a claim will only come gradually as a diversity of evidence accumulates.

Reluctance to accept the optimal signal detection theory framework stems largely from its mathematical complexity and resulting opaqueness to the research community. For an assertion about anthropogenically induced climate change to be convincing, it must be understandable by scientists in neighboring disciplines.

The works of N. Wiener and later treatises on space–time signal processing are neither familiar nor are they easy reading for most physical scientists. One of our goals in this paper is to rederive the optimal detection procedure in a simple framework that will make it more accessible and intuitive. As so often happens after a laborious and rigorous derivation or proof, one finds a simple geometrical interpretation of the process that is more intuitive and leads to a greater understanding. This is the case here, as will be seen in later sections.

The most recent development in the climate signal detection problem is the inclusion of several possible sources of climate forcing. It has become apparent that in the case of anthropogenic climate forcing, one must include both the forcings due to greenhouse gases and to anthropogenic sulfate aerosols (Santer et al. 1995; Santer et al. 1996; Hegerl et al. 1997). Recent research using the GFDL climate models has found that the surface temperature response to these two climate forcings is linearly additive in both space and time (Haywood et al. 1997; Ramaswamy and Chen 1997). There are at least two other significant climate forcings to be considered (IPCC 1996). These are the naturally occurring climate forcings due to volcanic stratospheric aerosols and solar variability (10–11-yr solar cycle). We consider these four climate forcings to be deterministic because we have proxy information about the sources of these forcings and so can include them in our ensemble of simulations.

The signal detection problem is made more difficult

---

*Corresponding author address:* Dr. Mark J. Stevens, UCAR GPS/MET, P.O. Box 3000, Boulder, CO 80307-3000.  
E-mail: stevens@ucar.edu

by the inclusion of multiple deterministic signals amid a background of natural variability. To detect the climate signal-of-interest, a filter must be constructed, which will not only separate out the signal-of-interest from the background noise but also be able to discriminate the signal-of-interest from the other climate signals present. All of the climate signals considered here have some overlap in their space–time patterns or “fingerprints.” Any filter that allows the signal-of-interest to be passed through will also allow some component of the other signals to be passed through. This additional contaminating signal will be mistakenly identified as part of the signal that is being searched for. The approach we use to avoid this problem is to project out the component of the signal-of-interest, which is perpendicular to the signal of the other three combined climate forcings. This perpendicular component of the signal-of-interest is then used in the construction of the filter. The disadvantage of this approach is that the perpendicular component of the signal will be smaller than the total signal. This will result in some decrease in the signal-to-noise ratio and hence raise the threshold of detection.

As with all procedures of this kind, models play an important role. A model is used to generate a space–time waveform for each of the climate signals or combination of climate signals. Additional models are also used to calculate the optimal weights used in the construction of the optimal filter. The use of four different climate models allows us to make separate estimations of the signal strength. The different estimates are fairly consistent with one another indicating that the procedure is robust with respect to choice of climate model.

A byproduct of our procedure is the theoretical signal-to-noise ratio  $\gamma$ , which allows an assessment to be made of the performance of the procedure without use of the actual observed data. The square of the signal-to-noise ratio  $\gamma^2$  can be written as a sum over spectral contributions from different space–time modes in the problem. By looking at the composition of this distribution, we can assess where we are gaining the most performance and where there is negligible contribution.

The main findings are that the volcanic aerosol and greenhouse gas signals have large  $\gamma$  and are highly significant, the anthropogenic aerosol signal has a small  $\gamma$  but is also very significant, and the solar variability signal has both a small  $\gamma$  and very low significance. The total signal obtained when all four climate forcings are combined is so significant ( $\approx 99.9\%$  confidence) as to virtually rule out the possibility of its occurrence arising from the noise of natural variability.

## 2. Combining two or more estimators

Consider estimating the temperature of a reservoir by using two measuring devices. Let the individual estimators be of the form  $\hat{T}_i = T + \epsilon_i$ , where  $T$  is the true temperature of the reservoir and  $\epsilon_i$  is the random error in an individual estimate. The errors are assumed to have

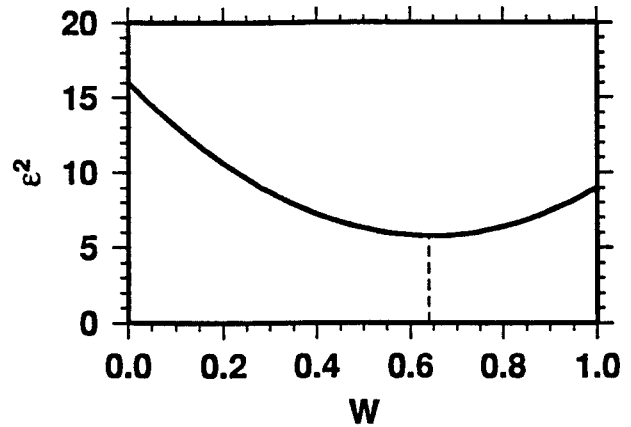


FIG. 1. Mean-square-error ( $\epsilon^2$ ) as a function of the weight ( $W$ ) for the case of  $\sigma_1 = 3$  and  $\sigma_2 = 4$ .  $W_{\text{opt}} = 0.64$ ,  $\epsilon_{\text{opt}}^2 = 5.76$ .

a Gaussian distribution with a mean of zero and to be independent (uncorrelated). These properties of the errors allow us to write  $\langle \epsilon_i \rangle = 0$  and  $\langle \epsilon_i \epsilon_j \rangle = \sigma_i^2 \delta_{ij}$ , where  $\langle \cdot \rangle$  represents the ensemble average, and  $\sigma_i^2$  are the error variances of the individual measuring devices. The individual estimator  $\hat{T}_i$  is unbiased, since  $\langle \hat{T}_i \rangle = \langle T + \epsilon_i \rangle = T + \langle \epsilon_i \rangle = T$ .

An estimate of the temperature of the reservoir can be written as a linear combination of the two independent estimates from the two measuring devices:

$$\hat{T} = W\hat{T}_1 + (1 - W)\hat{T}_2, \quad (1)$$

where  $W$  is an adjustable weight. The mean-square-error (mse) of the estimate is given by

$$\epsilon^2 = \langle (\hat{T} - T)^2 \rangle. \quad (2)$$

This can be expanded to give an equation, which is quadratic in  $W$ :

$$\epsilon^2 = (\sigma_1^2 + \sigma_2^2)W^2 - 2\sigma_2^2W + \sigma_2^2. \quad (3)$$

Figure 1 is a plot of the mse as a function of the weight  $W$  for the case of  $\sigma_1 = 3$  and  $\sigma_2 = 4$ . The figure shows that the value of the mse is somewhat insensitive to the exact value of  $W$  as long as  $W$  is near the minimum of the mse.

The optimal weight  $W$ , which will minimize the mse, can be found from the above equation by setting the derivative  $\partial \epsilon^2 / \partial W = 0$  and solving for  $W$ :

$$W_{\text{opt}} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \quad (4)$$

Substituting this result into (3) produces the optimal value of the mse:

$$\epsilon_{\text{opt}}^2 = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}. \quad (5)$$

We define

$$\eta^2 = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \tag{6}$$

so that  $\varepsilon_{\text{opt}}^2 = 1/\eta^2$ . By substituting (4) into (1) we get the important result

$$\hat{T}_{\text{opt}} = \frac{1}{\eta^2} \left( \frac{\hat{T}_1}{\sigma_1^2} + \frac{\hat{T}_2}{\sigma_2^2} \right). \tag{7}$$

These results can be generalized for  $K$  independent unbiased estimators as

$$\varepsilon_{\text{opt}}^2 = \frac{1}{\eta^2} = \frac{1}{\sum_{k=1}^K \frac{1}{\sigma_k^2}} \tag{8}$$

and

$$\hat{T}_{\text{opt}} = \frac{1}{\eta^2} \sum_{k=1}^K \frac{\hat{T}_k}{\sigma_k^2}. \tag{9}$$

Equation (8) shows that adding another unbiased measuring device always decreases the mse no matter how poor its quality (how large its error variance). This simple derivation of optimal weighting of independent estimators is familiar to many atmospheric scientists. The result is very intuitive. We simply weight each estimator inversely according to its individual error variance.

The derivation presented above required that the individual errors be uncorrelated with one another. If this were not so, the coordinate axes could simply be rotated to the principal axes of the error or noise covariance matrix. Then the entire formalism goes through as before except in the rotated coordinate system. In climatology, this is the transformation to the Empirical Orthogonal Function (EOF) basis set.

In two dimensions, this is easily spelled out explicitly. Let the covariance matrix of the noise be given by  $\mathbf{C}_{ij} = \langle \varepsilon_i \varepsilon_j \rangle$ . Taking the noise to be distributed bivariate normally, the contours of equal probability of occurrence of pairs of values of  $(\varepsilon_1, \varepsilon_2)$  are given by (Thiébaux 1994)

$$\frac{T_1^2}{\sigma_1^2} - 2\rho \frac{T_1 T_2}{\sigma_1 \sigma_2} + \frac{T_2^2}{\sigma_2^2} = \text{constant}, \tag{10}$$

which is an ellipse in the  $(T_1, T_2)$  plane, and where  $\rho$  is the correlation coefficient ( $= \langle \varepsilon_1 \varepsilon_2 \rangle \sigma_1^{-1} \sigma_2^{-1}$ ). In the two-dimensional case, we can find an angle through which to rotate the coordinate axes so that the principal axes of the ellipse will coincide with the new coordinate axes  $(T'_1, T'_2)$ :

$$\theta_{\text{rot}} = \frac{1}{2} \arctan \left( \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2} \right). \tag{11}$$

In the new coordinate system,  $T'_1$  and  $T'_2$  are uncorrelated. In the case of  $K$  dimensions, the figure is an ellipsoid in the  $K$  dimensional space, and a simple length preserving rotation can also be used to find the appropriate coordinate system.

### 3. The fingerprint and optimal estimators

#### a. The state space

Consider a  $K$ -dimensional vector space, which is spanned by a set of orthonormal basis vectors  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_K\}$ . In this state space, the basis vectors lie along the principal axes of a  $K$ -dimensional ellipsoid, which defines the Gaussian probability density function. In other words, the noise covariance matrix is diagonal in this state space and the basis vectors are the orthonormal eigenvectors of the covariance matrix. Any Gaussian random variable can be represented by its statistically independent components, which lie along the direction of the basis vectors.

In this state space we can define a noise vector as

$$\mathbf{N} = n_1 \mathbf{e}_1 + n_2 \mathbf{e}_2 + \dots + n_K \mathbf{e}_K, \tag{12}$$

where the  $n_1, n_2, \dots, n_K$  are the random noise components, which will differ from one realization to another. Similarly, we can define a signal vector as

$$\mathbf{S} = s_1 \mathbf{e}_1 + s_2 \mathbf{e}_2 + \dots + s_K \mathbf{e}_K. \tag{13}$$

The unit vector in the direction of the signal is written as

$$\mathbf{e}_s \equiv \frac{\mathbf{S}}{S} = \left( \frac{s_1}{S} \right) \mathbf{e}_1 + \left( \frac{s_2}{S} \right) \mathbf{e}_2 + \dots + \left( \frac{s_K}{S} \right) \mathbf{e}_K, \tag{14}$$

where the magnitude of the signal vector is given by

$$S = (s_1^2 + s_2^2 + \dots + s_K^2)^{1/2}. \tag{15}$$

Since this is a vector space, the usual vector operations are defined. We also assume that the data can be written as a linear sum of the signal and noise. We can then write a vector for the data:

$$\begin{aligned} \mathbf{D} &= (s_1 + n_1) \mathbf{e}_1 + (s_2 + n_2) \mathbf{e}_2 + \dots \\ &+ (s_K + n_K) \mathbf{e}_K. \end{aligned} \tag{16}$$

As an example of the vector dot product, we form  $\mathbf{e}_1 \cdot \mathbf{D}$ :

$$\begin{aligned} \mathbf{e}_1 \cdot \mathbf{D} &= \mathbf{e}_1 \cdot [(s_1 + n_1) \mathbf{e}_1 + (s_2 + n_2) \mathbf{e}_2 + \dots + (s_K + n_K) \mathbf{e}_K] \\ &= s_1 + n_1, \end{aligned} \tag{17}$$

which is an unbiased estimator of  $s_1$ .

For the climate signal problem, we need to define what we mean by the state space. Consider three stations whose temperature readings are  $T_1, T_2$ , and  $T_3$ . These readings form the components of a three-dimensional vector, and they define the state of that array of stations at a particular time. The variations will consist of a sum of forced and free contributions. In general, the natural variability part of the temperatures of the three will be correlated so that it will be necessary to rotate the coordinate system to the EOF (principal component) basis to use optimal estimation methods efficiently. We could consider the evolution of this three-vector in time. Its tip will move in the space due to the influence of climate

noise and deterministic forcings. In our treatment, we go a step farther and use a spectral representation in time. Since the noise is assumed stationary in time, it is appropriate to use a Fourier basis set. In this case, for each Fourier frequency component, there will be a set of spatial EOFs. For example, the three stations may have different spatial patterns (correlation-free linear combinations) for each frequency. The state space for the climate problem then will consist of the Fourier components (sine and cosine) with each having a string of spatial EOFs attached. In the climate problem, which we will take up later in this paper, we will choose a frequency band containing eight discrete frequencies and 36 spatial boxes. This gives a total of  $8 \times 2 \times 36 = 576$  EOF components. In the Fourier representation, the state vector does not move in time but is stationary in an expanded space whose dimension depends on the frequency band chosen.

### b. Fingerprint estimator of $S$

A common problem in climatology is that we know the shape of the signal waveform but not its amplitude. In other words, we know the direction of the signal vector (indicated by the unit vector  $\mathbf{e}_S$ ) and want to find its magnitude  $S$ . An unbiased estimator of  $S$  is the scalar product  $\mathbf{e}_S \cdot \mathbf{D}$ . We show this for the two-dimensional case:

$$\langle \hat{S}_{\text{FP}} \rangle = \langle \mathbf{e}_S \cdot \mathbf{D} \rangle = \left( \frac{s_1}{S} \mathbf{e}_1 + \frac{s_2}{S} \mathbf{e}_2 \right) \cdot (d_1 \mathbf{e}_1 + d_2 \mathbf{e}_2), \quad (18)$$

or

$$\langle \hat{S}_{\text{FP}} \rangle = \frac{s_1^2 + s_2^2}{S} = \frac{S^2}{S} = S, \quad (19)$$

since  $\langle d_k \rangle = \langle s_k + n_k \rangle = s_k$  and where the subscript FP indicates ‘‘fingerprint.’’ In other words, we project the data vector  $\mathbf{D}$  along the direction of the signal vector  $\mathbf{S}$ . This fingerprint estimator has an mse given by

$$\epsilon_{\text{FP}}^2 = \left( \frac{s_1}{S} \right)^2 \sigma_1^2 + \left( \frac{s_2}{S} \right)^2 \sigma_2^2. \quad (20)$$

The above fingerprint estimator ( $\mathbf{e}_S \cdot \mathbf{D}$ ) can also be written as

$$\hat{S}_{\text{FP}} = (\mathbf{e}_S \cdot \mathbf{e}_1)(\mathbf{e}_1 \cdot \mathbf{D}) + (\mathbf{e}_S \cdot \mathbf{e}_2)(\mathbf{e}_2 \cdot \mathbf{D}). \quad (21)$$

This can be generalized to  $K$  dimensions and rewritten as

$$\hat{S}_{\text{FP}} = \left\{ \sum_{k=1}^K (\mathbf{e}_S \cdot \mathbf{e}_k) \mathbf{e}_k \right\} \cdot \mathbf{D}. \quad (22)$$

The fingerprint method is very easy to implement and has attracted some users. On the other hand, it does not take advantage of the fact that the individual variances in the mse might be quite different. Hence, we might want to weigh the information from the component directions differently to take advantage of the fact that

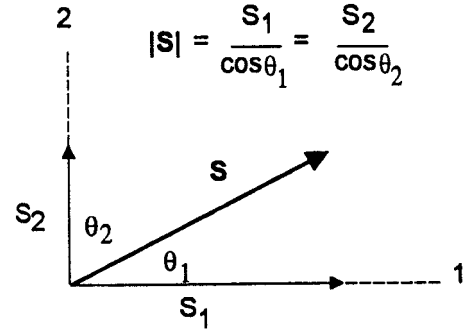


FIG. 2. Two independent estimators of the length of  $\mathbf{S}$  from its components and direction cosines.

one component might be more accurately estimated than another. The way to do this is presented in the next subsection.

### c. Optimal estimator of $S$

Consider that we have a signal vector  $\mathbf{S}$  and a data vector  $\mathbf{D}$ , both with  $K$  components. For the  $k$ th component of  $\mathbf{S}$  we can write

$$\mathbf{e}_S \cdot \mathbf{e}_k = \frac{s_k}{S} \quad (23)$$

or

$$S = \frac{s_k}{\mathbf{e}_S \cdot \mathbf{e}_k}. \quad (24)$$

For a two-dimensional example of this, refer to Fig. 2. Similarly, we can write for the  $k$ th component of  $\mathbf{D}$

$$\mathbf{e}_k \cdot \mathbf{D} = d_k = s_k + n_k. \quad (25)$$

The  $k$ th unbiased estimator of  $S$  can then be written as

$$\hat{S}_k = \frac{\mathbf{e}_k \cdot \mathbf{D}}{\mathbf{e}_S \cdot \mathbf{e}_k}, \quad (26)$$

since

$$\langle \hat{S}_k \rangle = \frac{\langle \mathbf{e}_k \cdot \mathbf{D} \rangle}{\mathbf{e}_S \cdot \mathbf{e}_k} = S. \quad (27)$$

We have  $K$  statistically independent unbiased estimators of  $S$ , one for each component direction. The error variance of the  $k$ th unbiased estimator can be shown to be

$$\text{Var}(\hat{S}_k) = \frac{\sigma_k^2}{(\mathbf{e}_S \cdot \mathbf{e}_k)^2}, \quad (28)$$

where  $\langle n_k^2 \rangle = \sigma_k^2$  is the noise variance in the  $k$ th component. We want to linearly combine these  $K$  estimators to give the optimal unbiased estimate of  $S$ . This is the same problem as trying to estimate the temperature of the reservoir in the previous section. In analogy with (9) we can write

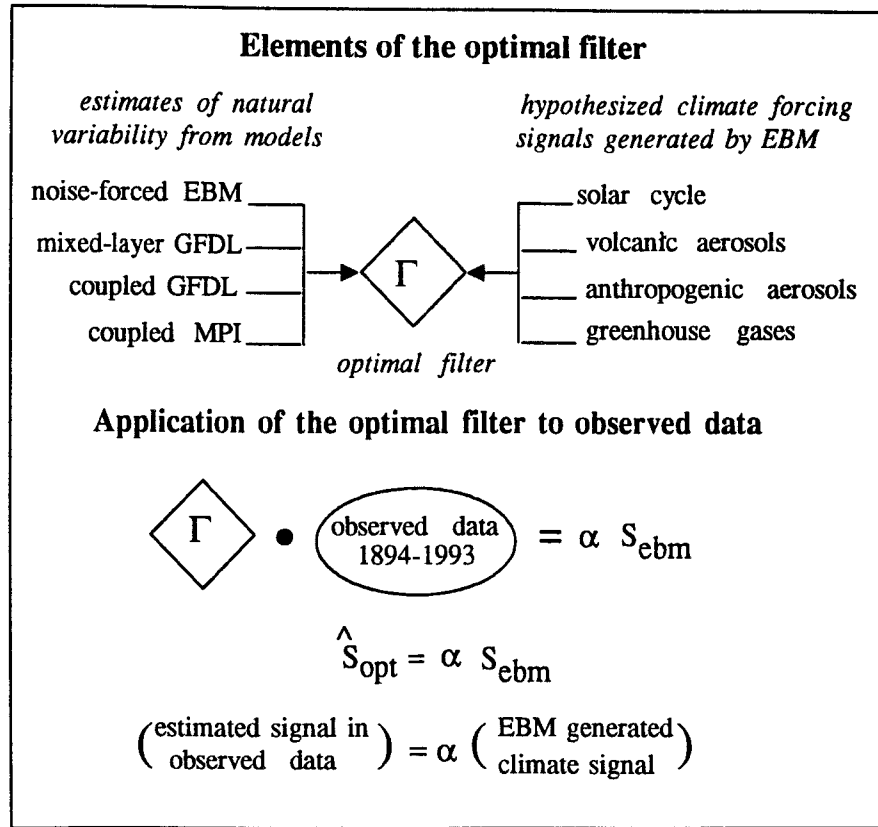


FIG. 3. Schematic diagram showing the elements in the construction of the optimal filter and its application to the observed data.

$$\hat{S}_{opt} = \frac{1}{\eta^2} \sum_{k=1}^K W_k \hat{S}_k, \quad (29)$$

where the  $W_k$  are the optimal weights given by

$$W_k = \frac{1}{\text{Var}(\hat{S}_k)}, \quad (30)$$

and the estimators  $\hat{S}_k$  are given by (26). Hence, the optimal estimator of  $S$  can be written as

$$\hat{S}_{opt} = \frac{1}{\eta^2} \sum_{k=1}^K \frac{(\mathbf{e}_s \cdot \mathbf{e}_k)(\mathbf{e}_k \cdot \mathbf{D})}{\sigma_k^2} \quad (31)$$

with

$$\eta^2 = \sum_{k=1}^K \frac{(\mathbf{e}_s \cdot \mathbf{e}_k)^2}{\sigma_k^2}. \quad (32)$$

Equation (31) can be rewritten as

$$\hat{S}_{opt} = \left\{ \frac{1}{\eta^2} \sum_{k=1}^K \frac{(\mathbf{e}_s \cdot \mathbf{e}_k) \mathbf{e}_k}{\sigma_k^2} \right\} \cdot \mathbf{D} \equiv \mathbf{\Gamma} \cdot \mathbf{D}, \quad (33)$$

where  $\mathbf{\Gamma}$  is defined as the optimal filter. In the last expression for  $\hat{S}_{opt}$ , we show the data vector  $\mathbf{D}$  factored out to emphasize that the procedure is a projection operation on the data vector  $\mathbf{D}$ ; hence, the term *filter*.

Compare this equation with (22) for the fingerprint estimator.

To emphasize the connection between this derivation and the derivation presented in North et al. (1995), and that appears in Stevens and North (1996), we substitute (23) into (31) and (32), and (25) into (31) and after some rearranging get

$$\hat{S}_{opt} = S \left( \frac{1}{\gamma^2} \sum_{k=1}^K \frac{s_k d_k}{\sigma_k^2} \right) \equiv S \alpha, \quad (34)$$

where  $\alpha$  is the scaling factor (Stevens and North 1996), and  $\gamma^2$  is the square of the theoretical signal-to-noise ratio:

$$\gamma^2 \equiv \sum_{k=1}^K \frac{s_k^2}{\sigma_k^2} = S^2 \eta^2. \quad (35)$$

The scaling factor  $\alpha$  is the ratio of the estimated magnitude of the signal in the data to the magnitude of the model signal. Figure 3 is a schematic diagram, which summarizes these ideas. In Stevens and North (1996), the summation over  $k$  in (34) and (35) refers to summation over the eigenmodes, while the noise variance  $\sigma_k^2$  is replaced by the eigenvalue  $\lambda_k$  corresponding to the  $k$ th eigenmode.

Let us recall our assumptions. First and foremost, we



TABLE 1. Climate signal vector notation.

Climate signal	Vector	Unit vector
Greenhouse gases	<b>G</b>	$\mathbf{e}_G \equiv \mathbf{G}/G$
Anthropogenic aerosols	<b>A</b>	$\mathbf{e}_A \equiv \mathbf{A}/A$
Volcanic aerosols	<b>V</b>	$\mathbf{e}_V \equiv \mathbf{V}/V$
Solar variability	<b>S</b>	$\mathbf{e}_S \equiv \mathbf{S}/S$

assumed the linear additivity of the signal and the noise. We have used the principal component directions to formulate the problem from the beginning. That is, we chose the coordinate axes to be the principal axes of the covariance ellipsoid of the noise vector. We had to assume knowledge of the direction of the signal waveform. Our job is to estimate the strength of the signal in the data given this information and one realization of the data. The quantity  $\gamma = S\eta$  is a good a priori measure of the quality of the procedure, since it is the theoretical signal-to-noise ratio. We can use  $\gamma^2$  as computed with models to tell us which vector components are most important in the estimation problem without invoking the observed data. This is very important, since we can use our climate models to condition our choice of the subspace within which we can make a reliable estimation of the signal strength without involving the data itself (which might be considered cheating).

#### 4. Interference by other signals

In the climate problem, there are many deterministic climate signals present in the observed data along with the natural variability (climate noise). For example, there are the greenhouse gas signal **G**, the volcanic signal (stratospheric aerosols) **V**, the anthropogenic sulfate aerosol signal **A**, and the solar cycle signal **S** (see Table 1 for notation). Hereafter, **S** will refer to the solar cycle signal, not to some general signal vector. Any of these signal vectors can be described in relation to any of the others. For example, consider the volcanic signal **V** as composed of a part that is parallel to the solar cycle vector **S** and a part that is perpendicular to **S**:

$$\mathbf{V} = \mathbf{V}_{\parallel S} + \mathbf{V}_{\perp S}. \quad (36)$$

If the angle between the **V** and **S** vectors is  $\theta_{V,S}$ , then

TABLE 2. Angles between possible pairs of the signal vectors.

Vector pair	Angle (deg)
<b>S</b> · <b>G</b>	77.9
<b>S</b> · <b>V</b>	88.0
<b>S</b> · <b>A</b>	101.0
<b>V</b> · <b>A</b>	84.2
<b>V</b> · <b>G</b>	93.8
<b>G</b> · <b>A</b>	153.3
<b>S</b> ·( <b>VAG</b> )	80.9
<b>V</b> ·( <b>GAS</b> )	90.9
<b>G</b> ·( <b>VAS</b> )	116.9
<b>A</b> ·( <b>SVG</b> )	133.6

TABLE 3. Lengths of the signal vectors as calculated by the EBM.

Vector	Magnitude
<b>S</b>	6.98
<b>A</b>	13.11
<b>V</b>	20.24
<b>G</b>	28.56
<b>SVG</b>	28.49

the component of **V**, which is perpendicular to **S**, is given by

$$V_{\perp S} = V \sin \theta_{V,S}, \quad (37)$$

and the component which is parallel to **S** is given by

$$V_{\parallel S} = V \cos \theta_{V,S}. \quad (38)$$

As the angle between the vectors approaches  $90^\circ$ , the perpendicular component approaches  $V$ , and the parallel component approaches zero. However, if the angle  $\theta_{V,S}$  is not close to  $90^\circ$ , then the volcanic signal will definitely have a component, which lies along the direction of the solar signal (given by the unit vector  $\mathbf{e}_S = \mathbf{S}/S$ ). This could present a problem if the objective is to detect a weak solar signal and the volcanic signal is strong. An optimal filter that has been constructed to allow the solar signal to pass through will also allow to pass through the component of the volcanic signal, which lies along the direction of the solar signal. The volcanic signal can then be considered a source of interference in the search for the solar signal. The obvious way around this problem is to use the part of the solar signal that is perpendicular to the volcanic signal when constructing the optimal filter. The disadvantage here is that for angles that are far from  $90^\circ$ , the perpendicular component could be much smaller than the total magnitude of the signal. This might only present a problem when the signal-of-interest has a small signal-to-noise ratio. This idea of interference between signals can also be

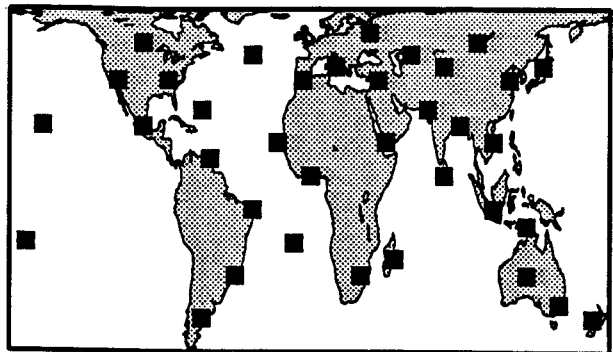


FIG. 4. Location of the 36 detection boxes. There are 20 boxes in the Tropics and 16 in the extratropics, 24 in the Northern and 12 in the Southern Hemisphere. Each of the 36  $10^\circ \times 10^\circ$  detection boxes is comprised of four  $5^\circ \times 5^\circ$  boxes from the Jones dataset, each of which has 1200 months of data (1894–1993). Boxes were chosen based on where there was sufficient data, spatial sampling was maximized, and correlation between the boxes was minimized.

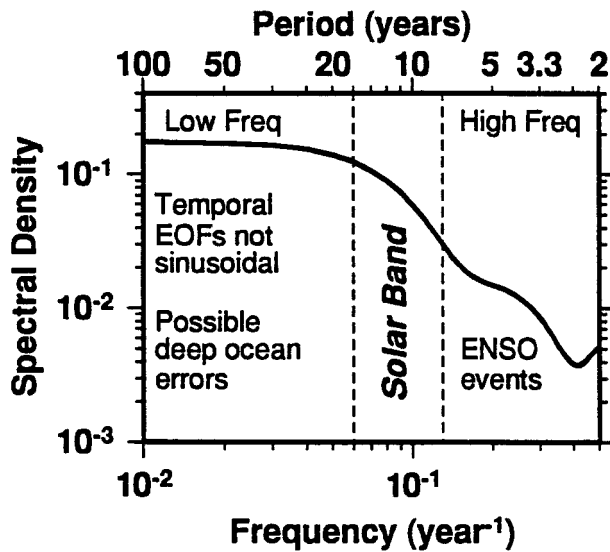


FIG. 5. Smoothed power spectra of observed global annual mean surface temperature anomalies and the solar cycle frequency band (periods from 16.67 to 7.69 yr).

applied when there are multiple signals involved. Each signal might be a source of contamination in the search for one particular signal.

Table 2 lists the separation angles between the possible pairs of signals produced by the four climate forcings used in this study. Also listed are the angles between the individual signals and the signal resulting from the combination of the other three climate forcings. For example, the signal **VAG** is the signal that results when the forcings due to volcanic aerosols, greenhouse gases, and anthropogenic aerosols are combined. As can be seen in Table 2, the angle between the solar signal **S** and the signal **VAG** is  $80.9^\circ$ . Thus, we might expect some contamination from the signal **VAG** when looking for **S** in the data. In Table 3, the magnitudes of the four signal vectors are listed along with the magnitude of the signal resulting from combining all four climate forcings. Notice that both **V** and **G** are much larger than **S**. The magnitude of **SVGA** is smaller than that of **G** because the forcings **V** and **A** act to cool the surface of the earth, while **G** acts to warm it. Thus, the vector sum  $\mathbf{S} + \mathbf{V} + \mathbf{G} + \mathbf{A}$  involves some cancellation of components.

### 5. Errors in the optimal filter

As can be seen from (32) and (33), the construction of the optimal filter  $\mathbf{\Gamma}$  requires climate models to generate a signal (for  $\mathbf{e}_s$ ) and the background climate noise (for  $\sigma_k^2$ ). Since climate models are not perfect representations of reality, they will introduce errors into the detection problem. We now examine some possible errors in the construction of the optimal filter.

All filters have an inherent type of error variance that we call the filter “pass-thru” error. This error arises

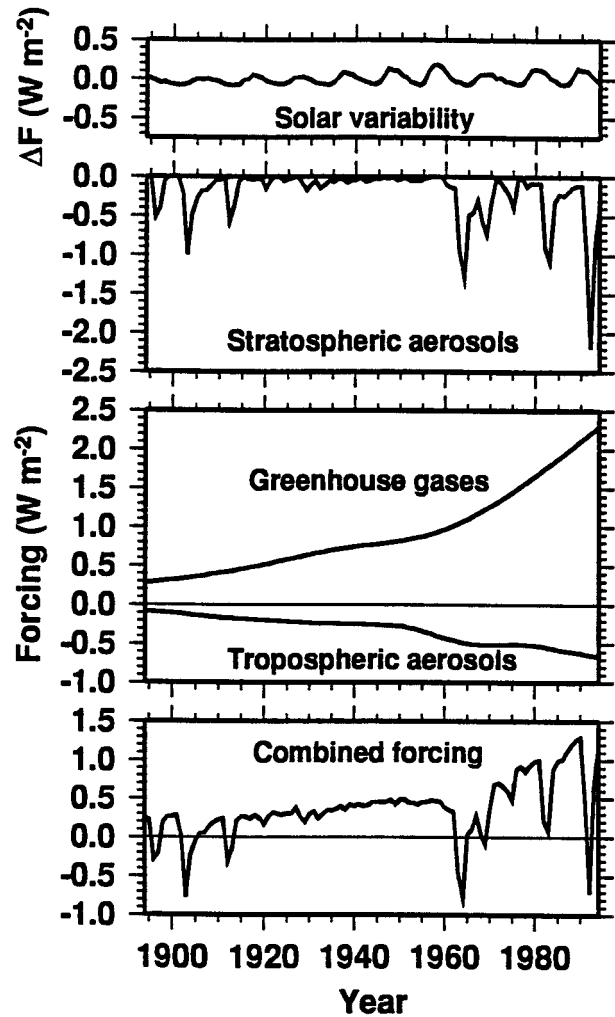


FIG. 6. Time series of global annual mean climate forcings ( $\text{W m}^{-2}$ ) used to generate climate signals in the EBM. Note that the vertical scaling of the plots is the same for each panel.

from the fact that some component of the background noise will always be passed through the filter because by chance it looks like the signal. This error has a theoretical value  $\sigma_{\text{pass}}^2 = 1/\gamma^2$ . The larger the  $\gamma$  of the model signal used to construct the filter, the smaller the pass-thru error. This type of error will exist no matter how closely the models can approximate reality.

Consider the error variance involved in the use of imperfect models in constructing the filter. The first type of error results from using an incorrect fingerprint. In the present context, this means the unit vector  $\mathbf{e}_s$  has the wrong direction in the state space. An equivalent statement is that the direction cosines  $\mathbf{e}_s \cdot \mathbf{e}_k$  are incorrect. The single constraint on these direction cosines is that their squares must add up to unity. An incorrect fingerprint can lead to a bias in the estimation of the signal strength. We call this type of error the “model bias” error, or  $\sigma_{\text{bias}}^2$ .

Another type of error variance comes from the op-

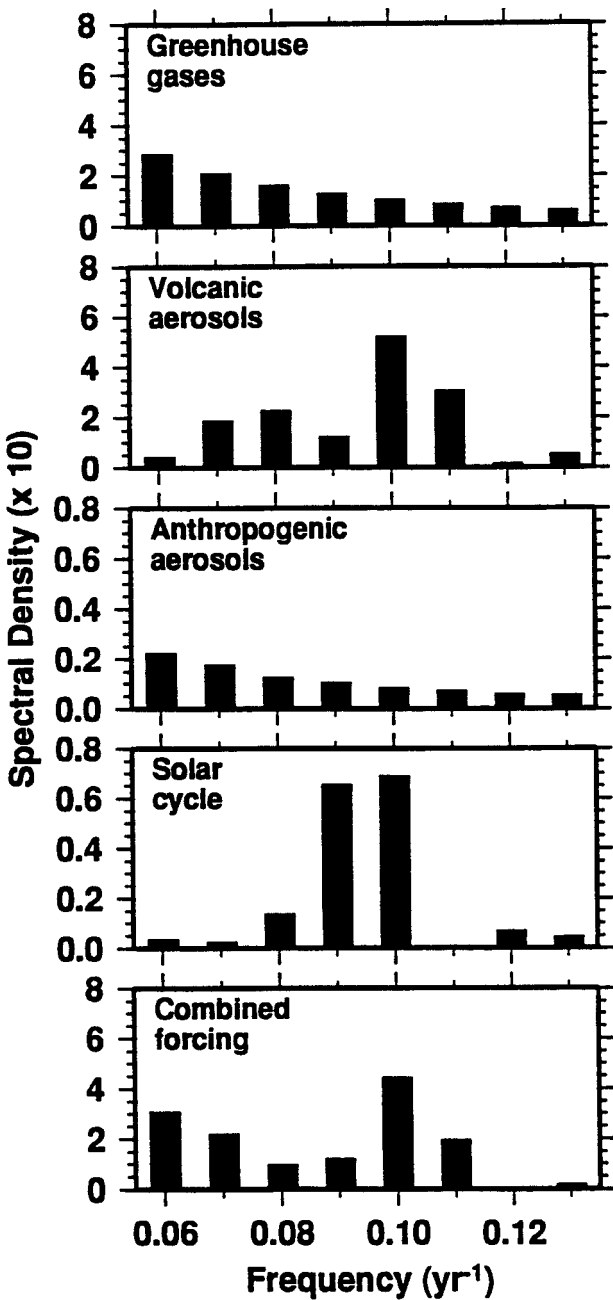


FIG. 7. Power spectra of the four individual global annual mean forcings and the combined global annual mean forcing. Notice the vertical scale varies by a factor of 10. The power at the frequency  $0.10 \text{ yr}^{-1}$  is dominated by the volcanic aerosol forcing.

timal weights as generated from models. This type of error is less egregious than error in the signal waveform. Since the estimator is composed of  $K$  independent estimators, which are assumed to be unbiased, the weighting does not introduce a bias. If the weights are erroneous, they can lead to a suboptimal estimator, and they can lead to an underestimation of the theoretical mean squared error ( $1/\eta^2$ ). It turns out that since the minimum

in the mse as a function of the weights is the minimum of a multidimensional parabolic surface (actually the intersection of this parabolic surface with the plane  $\sum_k W_k = 1$ ), the mse is not sensitive to the exact choice of the weights. We call this type of model error the “filter sampling” error,  $\sigma_{\text{samp}}^2$ , since it can arise from having a limited number of samples of model-generated noise from which to estimate the optimal weights. Our estimate of the sampling error is much reduced from the estimate in Stevens and North (1996), since we now truncate the number of eigenmodes used in the calculation of the optimal filter. This is done to limit the possible spurious contributions from the higher eigenmodes, which have small eigenvalues. Eigenvalues that contribute less than 1% of the total variance were not used (Kim 1996). The estimated total error variance can then be written as the sum

$$\sigma_{\text{total}}^2 \approx 1/\gamma^2 + \sigma_{\text{bias}}^2 + \sigma_{\text{samp}}^2. \quad (39)$$

Using our new estimate of  $\sigma_{\text{samp}}^2$  and our previous estimate of  $\sigma_{\text{bias}}^2$ , we can estimate the total error variance as

$$\sigma_{\text{total}}^2 \approx 1/\gamma^2 + 0.05. \quad (40)$$

Of course there are other sources of error that are difficult or impossible to estimate so this estimate of the total error should be considered as the minimum error estimate.

## 6. Application to 100 yr of temperature data

Stevens and North (1996) have analyzed 100 yr (1894–1993) of surface temperature data based upon  $36 \text{ } 10^\circ \times 10^\circ$  boxes distributed over the earth as shown in Fig. 4. The data is from the Jones dataset of monthly temperature anomalies (Jones and Briffa 1992). Each box results from averaging the data of four smaller  $5^\circ \times 5^\circ$  boxes. The boxes were chosen based on the continuity, representativeness, and distribution of the data. Each box has 1200 months of continuous data. In our analysis, we excluded the winter half-year of data in the extratropics and kept all 12 months of data in the Tropics as explained in Stevens and North (1996). In future work we intend to keep all months using a cyclostationary EOF basis set (Kim et al. 1996). The retained data form a 36-component multivariate discrete time series. In the absence of external forcing, the time series should be stationary as we have found using 1000-yr time series (control runs) from several coupled ocean–atmosphere GCMs.

By analyzing these long control runs, we were able to calculate the spatial EOFs for each Fourier frequency component (actually, we took the spatial factors to be the same over the narrow frequency band employed in the previous study as well as here). Since our original interest was in detecting the solar signal amid climate noise and other deterministic signals, we restrict our investigation to the “solar band,” which includes only the eight frequencies from  $0.06$  to  $0.13 \text{ yr}^{-1}$ . The cor-



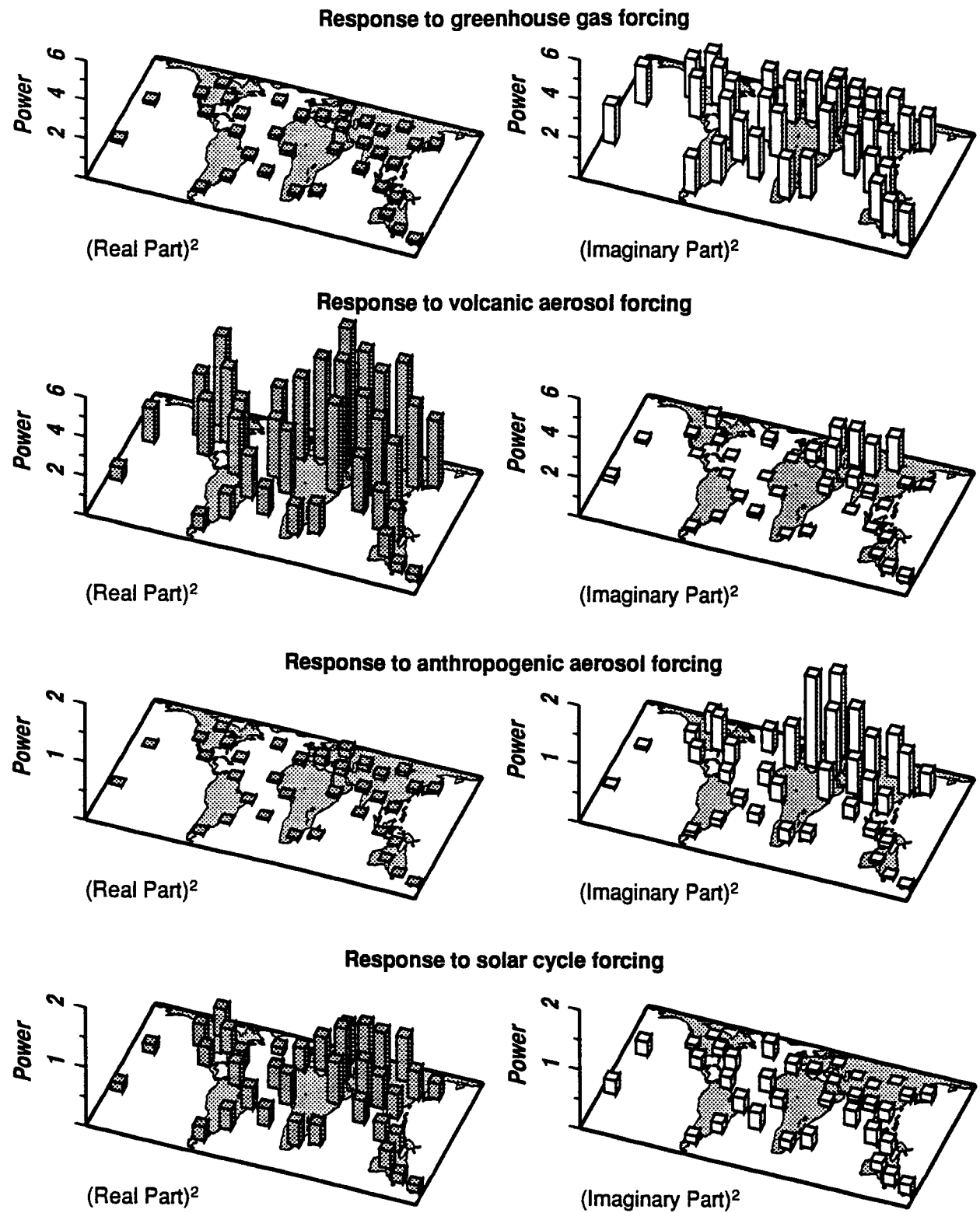


FIG. 8. Squares of the real and imaginary parts from the DFT of the EBM response to the four climate forcings at the frequency  $0.10 \text{ yr}^{-1}$  (10-yr period). Note the change in the vertical scale.

TABLE 4. Theoretical signal-to-noise ratio  $\gamma$ , and scaling factor  $\alpha$ , calculated using the component of the solar signal  $\mathbf{S}$ , perpendicular to the signal vector  $\mathbf{VAG}$ .

Region	EBM		GFDL/ml		GFDL/c		MPI	
	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$
Global	1.56	0.33	1.48	-0.09	1.55	0.47	2.44	0.30
Tropics	1.52	0.37	1.64	0.21	1.55	0.10	2.38	0.07
Extratropics	1.09	0.05	0.93	0.20	0.93	0.18	1.60	0.21
N. Hem.	1.33	0.47	1.33	-0.12	1.29	0.18	2.31	0.21
S. Hem.	1.30	0.10	1.38	0.59	1.37	-0.19	2.08	-0.11
W. Hem.	1.13	-0.26	1.45	-0.02	1.34	-0.42	1.94	0.11
E. Hem.	1.05	1.02	1.28	0.43	1.35	0.77	2.19	0.04
Mean	1.47	0.30	1.36	0.17	1.34	0.16	2.13	0.12

responding timescales range from 16.67 to 7.69 yr. The band chosen here is purposely designed to exclude as much ENSO activity as possible ( $\text{freq} > 0.13 \text{ yr}^{-1}$ ) because the climate models being used do not simulate ENSO very faithfully, if at all. In our previous study (Stevens and North 1996), we retained the high-frequency  $0.14 \text{ yr}^{-1}$  harmonic. An additional advantage offered by the low-frequency cutoff is that the temporal eigenfunctions for the 100-yr segment are closer to sines and cosines if the frequency is high compared to the  $(\text{segment length})^{-1}$  (see the appendix of North and Kim 1995). See Fig. 5 for a summary of these considerations. A final important argument for using only the narrow band of frequencies in our study is that present coupled ocean-atmosphere GCMs may not be able to produce the low-frequency spectrum of fluctuations with any reliability. In the band of interest here, even a mixed-layer model appears to be adequate. Fluctuations in this band from the coupled GFDL model appear to be very similar to those in the corresponding mixed-layer model (Manabe and Stouffer 1996).

We compute the signals  $\mathbf{S}$ ,  $\mathbf{V}$ ,  $\mathbf{G}$ , and  $\mathbf{A}$  using an energy balance model described in Stevens (1997). Stevens (1997) also describes how the individual climate forcings are specified in both time and space for use in the EBM. Figure 6 shows the global annual means of the climate forcings used in the EBM. Each frequency in the solar band has a real (cosine) and imaginary (sine) phase component. Combining the 36 spatial components (boxes) with the eight frequencies and the two phases yields a state space of 576 dimensions ( $36 \times 8 \times 2$ ). Thus, the signal vectors  $\mathbf{S}$ ,  $\mathbf{V}$ ,  $\mathbf{G}$ , and  $\mathbf{A}$  each have 576 components in the state space.

## 7. Results

Our main goal is to find the optimal estimates of the strengths (magnitudes) of the signals  $\mathbf{S}$ ,  $\mathbf{V}$ ,  $\mathbf{A}$ , and  $\mathbf{G}$ . What we actually find is the ratio of the estimated strength of a signal in the data to the strength of the signal produced by the EBM (the scaling factor  $\alpha$ ). Figure 7 shows the Fourier spectral power (sums of sine squared and cosine squared components at each discrete frequency) for each of the four climate forcings in the

frequency band ranging from 0.06 to  $0.13 \text{ yr}^{-1}$ . Note the steady decrease in power with increasing frequency in the greenhouse gas and the anthropogenic aerosol spectra, indicative of the Fourier transform of a ramp forcing. Of course, these vectors are nearly oppositely directed in the 576-dimensional space (see Table 2). Note also that the units along the vertical axes differ by a factor of 10. Another interesting fact is the amount of power at frequencies of  $0.10 \text{ yr}^{-1}$  and  $0.11 \text{ yr}^{-1}$  in the forcing spectrum of the volcanic aerosols. We assert that this comes about from the chance spacings of volcanic eruptions over the last 100 yr (Fig. 6). In the combined forcing spectrum, the largest peak is at  $0.10 \text{ yr}^{-1}$ ; this is due to the volcanic aerosols, not to the solar variability, which is smaller by a factor of 10. This is indicative of the potential danger of not considering interfering deterministic signals. In Fig. 8 are shown the responses of the EBM to the individual climate forcings at a frequency of  $0.10 \text{ yr}^{-1}$ . The (real part)<sup>2</sup> refers to the real part (cosine) of the discrete Fourier transform (DFT), and (imaginary part)<sup>2</sup> refers to the imaginary (sine) part of the DFT. This phase information is included in the optimal filter and helps to discriminate one signal from another. Notice the large peaks in the response to the volcanic aerosol forcing and the fact that the greenhouse response is predominately in the imaginary part and fairly uniform over the 36 boxes.

In the following sections, we present the results for the component of the signal-of-interest, which is perpendicular to the signal produced by the sum of the other three climate forcings. In this way we form an estimate of the strength of each individual signal without interference from the other signals. In Tables 4 through 8 we show the results for  $\alpha$  and  $\gamma$  for each of four climate models used to calculate the optimal weights. EBM refers to a 10 000-yr control run from our noise-forced EBM; GFDL (Geophysical Fluid Dynamics Laboratory)/ml refers to a 1000-yr control run from the GFDL mixed-layer GCM; GFDL/c refers to a 1000-yr control run from the GFDL coupled ocean-atmosphere GCM; MPI refers to a 1000-yr control run from the Max Planck Institute coupled ECHAM1/LSG GCM. In the tables, the column labeled "Region" refers to the use of all 36 boxes (Global) or a subset of the 36 boxes (Fig.

TABLE 5. Theoretical signal-to-noise ratio  $\gamma$ , and scaling factor  $\alpha$ , calculated using the component of the volcanic signal  $\mathbf{V}$ , perpendicular to the signal vector  $\mathbf{GAS}$ .

Region	EBM		GFDL/ml		GFDL/c		MPI	
	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$
Global	4.19	1.33	4.10	0.93	4.24	1.20	6.75	1.04
Tropics	4.03	1.35	4.32	1.04	4.14	1.12	6.58	1.30
Extratropics	3.24	1.37	2.74	1.30	2.83	1.06	4.85	0.95
N. Hem.	3.86	1.27	3.88	1.06	3.82	0.88	6.79	0.98
S. Hem.	2.95	1.31	3.18	1.16	3.19	1.11	4.82	0.97
W. Hem.	3.15	1.34	3.94	1.14	3.77	1.26	5.26	0.88
E. Hem.	2.81	1.23	3.46	1.12	3.65	1.06	6.03	0.98
Mean	3.46	1.31	3.66	1.11	3.66	1.10	5.87	1.01

4). “Tropics” indicates that the 20 boxes in the Tropics (30°N–30°S) were used in the calculations. “Extratropics” indicates that the 16 boxes outside of the Tropics were used. “N. Hem.” indicates that the 24 boxes in the Northern Hemisphere were used, and similarly for S. Hem., W Hem., and E. Hem. (see Stevens and North 1996). In each of these cases, we computed a new set of optimal weights.

Figure 9 shows the global values for the scaling factor  $\alpha$  for the four models as well as the mean value of  $\alpha$  calculated from all the regions and models. The error bars represent the error estimate of  $\pm\sigma_{\text{total}}$  from (40). If there were no signals present in the data then we would expect  $\langle\alpha\rangle = 0.0$ . In Fig. 10, we compare results for  $\alpha$  calculated using the EBM global optimal weights in the optimal filters. The results for  $\alpha$  from the first column of Fig. 9 (EBM) are shown in Fig. 10 with horizontal error bars for each signal. The histograms of  $\alpha$  values in Fig. 10 were calculated using 100 control runs of climate noise generated by the EBM as data input to each optimal filter. Since these control runs contain no signal we should expect a Gaussian distribution of  $\alpha$  about a mean of zero. Notice that the sharpness of the distributions increases with increasing  $\gamma$ , as they should. Also notice that the  $\alpha$  values for the anthropogenic aerosol, volcanic aerosol, greenhouse gas, and combined forcing signals are well outside the distributions of the noise.

Finally, the results are summarized in Table 9, which shows the mean values of  $\alpha$  and  $\gamma$  when averaged over all regions and models. The error in the estimate of  $\alpha_{\text{mean}}$

is also given as the standard deviation of  $\alpha_{\text{mean}}$  from a value of zero. This standard deviation can be represented as a confidence level in the estimate of  $\alpha_{\text{mean}}$ , since we expect  $\alpha$  to be a positive number and so can use a one-tailed test of the hypothesis that the signal is in the observed data.

#### a. Solar cycle signal

Detecting the solar signal was our original goal. Stevens and North (1996) found a fairly robust solar signal when the other deterministic signals were ignored (EBM global:  $\gamma \approx 1.8$ ,  $\alpha \approx 1.0$ ). Taking the component, which is perpendicular to the individual signals, it is still barely detectable. For example,  $\gamma \approx 1.5$  and  $\alpha \approx 0.7$ , when we use the component perpendicular to the volcanic signal vector. Similar results are obtained when we look at the component perpendicular to the greenhouse gas and the anthropogenic vectors. But when we take the component, which is perpendicular to the signal produced by the sum of the other three forcings ( $\mathbf{VAG}$ ), we obtain the rather poor performance shown in Table 4. We can no longer claim to detect the solar cycle signal in the observed data with any degree of confidence.

We made another attempt to estimate the solar signal by taking the 50 yr of data, 1914–63, a period in which there was very little volcanic activity. When we considered the component perpendicular to the  $\mathbf{VAG}$  signal, we obtained the values  $\gamma \approx 0.8$  and  $\alpha \approx 0.7$ , using the EBM optimal weights. Clearly, some performance was gained by considering this quiet period, but a roughly

TABLE 6. Theoretical signal-to-noise ratio  $\gamma$ , and scaling factor  $\alpha$ , calculated using the component of the greenhouse gas signal  $\mathbf{G}$ , perpendicular to the signal vector  $\mathbf{VAS}$ .

Region	EBM		GFDL/ml		GFDL/c		MPI	
	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$
Global	3.28	1.93	3.06	1.54	3.23	1.87	5.03	1.75
Tropics	3.19	1.99	3.38	1.60	3.26	1.75	5.00	1.99
Extratropics	2.36	1.98	2.01	2.14	2.07	1.80	3.48	1.76
N. Hem.	2.86	1.94	2.83	1.78	2.74	1.53	4.84	1.65
S. Hem.	2.58	1.75	2.75	1.71	2.82	1.70	4.29	1.51
W. Hem.	2.49	2.03	3.13	1.73	2.92	1.96	4.13	1.46
E. Hem.	2.13	1.82	2.55	1.74	2.72	1.76	4.43	1.68
Mean	2.70	1.92	2.81	1.75	2.82	1.77	4.46	1.69

TABLE 7. Theoretical signal-to-noise ratio  $\gamma$ , and scaling factor  $\alpha$ , calculated using the component of the anthropogenic aerosol signal **A**, perpendicular to the signal vector **SVG**.

Region	EBM		GFDL/ml		GFDL/c		MPI	
	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$
Global	1.12	4.66	1.04	2.74	1.10	4.06	1.68	3.23
Topics	1.08	4.70	1.15	3.74	1.12	4.01	1.68	4.37
Extratropics	0.81	4.40	0.70	4.22	0.72	3.36	1.19	2.86
N. Hem.	0.99	4.40	0.97	3.31	0.94	2.55	1.62	2.95
S. Hem.	0.86	4.66	0.92	4.70	0.94	4.09	1.42	3.46
W. Hem.	0.85	4.40	1.03	4.09	0.97	4.40	1.36	3.17
E. Hem.	0.75	4.09	0.88	3.68	0.94	3.68	1.51	2.81
Mean	0.92	4.47	0.96	3.78	0.96	3.74	1.49	3.26

corresponding amount of signal was lost because of the shorter record.

#### b. Volcanic aerosol signal

The results for the volcanic signal are highly significant, as can be seen in Table 5. For the global case, all four models have a  $\gamma > 4$ . This means the signal strength is nonzero with a probability of greater than 99.99%. Its strength averages 1.12 for all the global data across the four models. We also computed  $\alpha$  for the volcanic signal perpendicular to the individual signals (details not shown in this paper). The  $\gamma$  was similarly large, and the value of  $\alpha$  was again just above unity.

#### c. Greenhouse gas signal

Table 6 shows that the results for the greenhouse gas signal are also highly significant. For the global case, all four models have a  $\gamma > 3$ . However, compared to the volcanic signal, the values for  $\alpha$  are much larger with an average of 1.77. This indicates that the greenhouse signal in the observed data is about 77% larger than the signal generated by the EBM. The large value of  $\alpha$  results in a very high level of significance, since  $\alpha$  is even farther from zero, and the  $\gamma$  is large. The average results for  $\alpha$  indicate a confidence level  $>99.99\%$ .

There may be several reasons why the signal generated by the EBM is too small. The greenhouse gas forcing used in the model (Stevens 1997) was calculated by Kiehl and Briegleb (1993) and represents the net

downward radiative forcing at the tropopause. Since the EBM is two-dimensional, it necessarily ignores potentially important processes in the vertical. The outgoing radiation is simply described by a linear relation with the surface temperature. For the EBM to show a similar response as found in the observed data, it might be necessary to increase the greenhouse gas forcing by about 77%. However, the statistical significance of  $\alpha$  values  $>1$  does not have much dependence on the model formulation. If the model forcing, or sensitivity, were increased so that  $\alpha \approx 1$ , the response (signal) would be larger. This would result in a proportional increase in  $\gamma$ , and the estimated error would decrease (since  $\sigma_{\text{total}}^2 \approx 1/\gamma^2 + 0.05$ ). The statistical significance of  $\alpha$  would only decrease slightly.

#### d. Anthropogenic aerosol signal

This is the weakest signal in terms of the  $\gamma$ , as can be seen in Table 7. This by itself would indicate a large uncertainty, since the pass-through error of the filter is equal to  $1/\gamma^2$ . Indeed, Table 9 shows that this signal has the largest error estimate. However, the global average value of  $\alpha$  is 3.67. So even with the largest error,  $\alpha$  is so far from zero that its significance is  $>99.9\%$ . Again, the reason that  $\alpha$  is so large is probably due to an underestimate of the forcing at the surface. One possible explanation is that we have only included the direct effect of the anthropogenic aerosols and that the indirect effect (involving cloud nuclei generation) is actually several times larger (Boucher and Lohmann 1995; Jones et al. 1994).

TABLE 8. Theoretical signal-to-noise ratio  $\gamma$ , and scaling factor  $\alpha$ , calculated using the combined signal vector **SVGA**.

Region	EBM		GFDL/ml		GFDL/c		MPI	
	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$	$\gamma$	$\alpha$
Global	8.33	0.82	7.25	0.64	7.82	0.73	11.28	0.82
Tropics	8.01	0.85	8.41	0.88	8.30	0.99	11.55	0.77
Extratropics	5.58	0.91	5.03	0.75	5.16	0.77	8.03	0.89
N. Hem.	6.61	0.76	6.27	0.67	5.91	0.71	9.81	0.67
S. Hem.	7.07	0.89	7.51	0.97	7.95	0.98	11.87	1.01
W. Hem.	6.39	0.92	7.60	0.62	6.97	0.67	9.76	0.66
E. Hem.	5.39	0.93	6.00	0.69	6.49	0.78	9.93	0.94
Mean	6.77	0.87	6.87	0.75	6.94	0.80	10.32	0.82

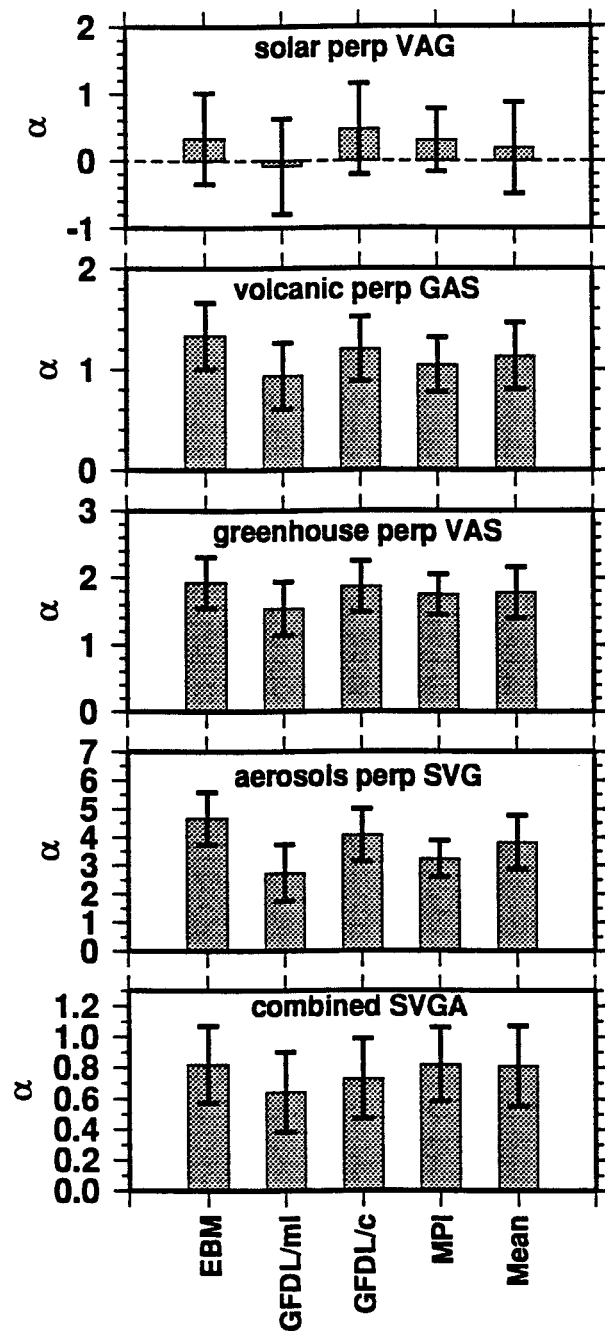


FIG. 9. Results from the four climate models for the scaling factor  $\alpha$  when all 36 detection boxes are used (global), and the mean value from averaging results from all the models and regions. The error bars are from the estimated total error.

*e. Total deterministic signal*

When all four climate forcings are used together to generate a signal, we get a very large  $\gamma$ . Table 8 shows that  $\gamma > 5$  for all regions and models. The mean value of  $\alpha$  from Table 9 is 0.81. Together these result in a very high confidence level of about 99.9%. The fact that  $\alpha$  is a little less than one indicates that the vector sum of all

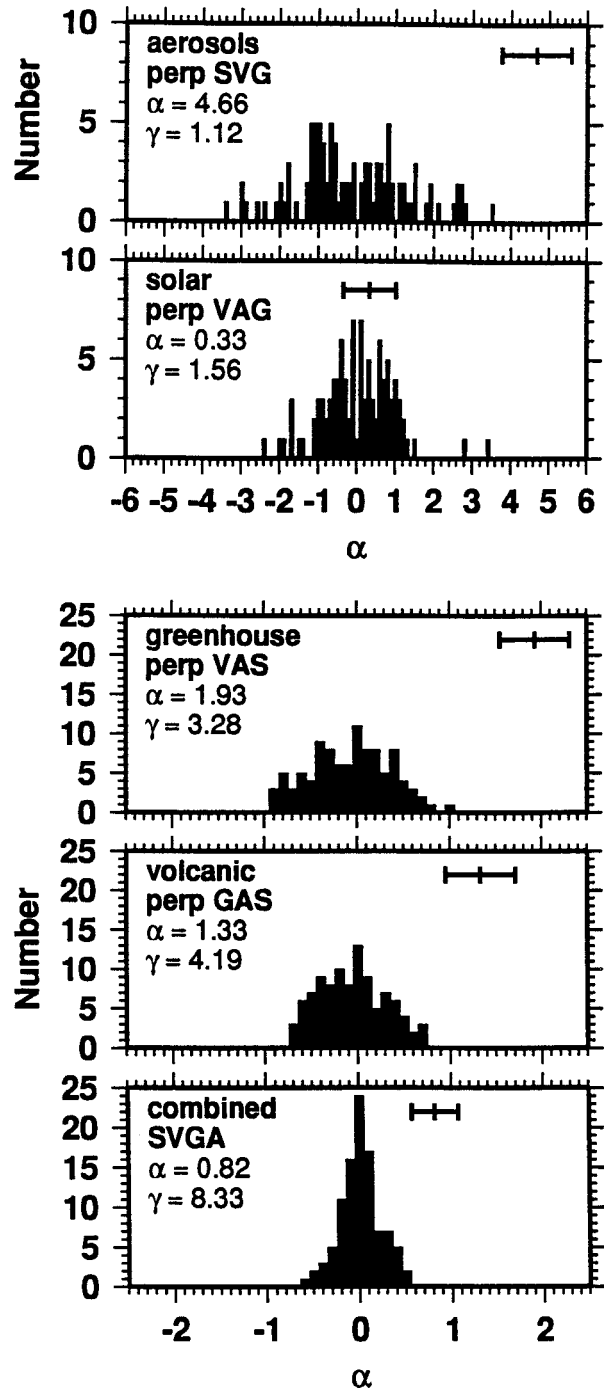


FIG. 10. Histograms of  $\alpha$  values calculated using 100 100-yr control runs from the noise-forced EBM as the data input to the optimal filter. The EBM global optimal weights were used. Also shown are the values of  $\alpha$  for the climate signals when the observed data is used as input to the optimal filter. The estimated total error is shown by the horizontal error bars.



TABLE 9. Summary of statistics for the five climate forcings. The columns  $\gamma_{\text{mean}}$  and  $\alpha_{\text{mean}}$  refer to the average values of  $\gamma$  and  $\alpha$  for all models and regions for each forcing (Fig. 9). The standard deviation of  $\alpha_{\text{mean}}$  from a mean of zero is given by  $\alpha_{\text{mean}}/\sigma_{\text{total}}$ . The last column refers to the confidence level using a one-tailed test.

Signal	$\gamma_{\text{mean}}$	$\alpha_{\text{mean}}$	$\sigma_{\text{total}}$	std dev	c.l. (%)
Solar	1.56	0.19	0.68	0.28	61.0
Volcanic	4.16	1.13	0.33	3.42	99.97
Greenhouse	3.20	1.78	0.38	4.68	>99.99
Aerosols	1.08	3.81	0.95	4.01	>99.99
Combined	7.73	0.81	0.26	3.12	99.91

the signals is approximately the same in the observed data as in the EBM. This must mean that there is some cancellation going on between the climate forcings. The greenhouse gas signal acts in the opposite sense of both the volcanic and anthropogenic aerosol forcing signals, so we might expect them to cancel each other to some extent.

#### f. Decomposition of $\gamma^2$

The above results for the total deterministic signal raises a question about where the contributions to  $\gamma^2$  are coming from. Is the large  $\gamma$  due to just the global mean? Consider the formula for the square of the signal-to-noise ratio:

$$\gamma^2 = \sum_n \frac{T_n^2}{\lambda_n}, \quad (41)$$

where  $T_n$  is the projection of the signal onto the  $n$ th EOF and  $\lambda_n$  is the eigenvalue corresponding to that EOF. The index  $n$  is really composed of two parts, the frequency component with real and imaginary phases and the spatial EOF index. We sum the two phase contributions and then display in a three-dimensional bar graph the contributions from the various terms in the last equation. Figure 11 show this decomposition of the total deterministic signal. Note that contributions are spread broadly across many spatial EOF modes. The main point of this analysis is that many EOF components contribute to the  $\gamma^2$ ; it is not simply the global average (first EOF).

## 8. Conclusions

In this paper we presented a derivation of the optimal filter, which is equivalent to earlier formulations, but conceptually much simpler because of its relation to the simple optimal weighting of independent estimators and the use of vector notation. A scheme was also developed for estimating individual signal strengths when they are embedded in a mixture of deterministic signals. The method was applied to a series of signals, which might be expected to appear in the observed data.

An important finding of this analysis is the occurrence in the climate-forcing spectrum of strong peaks at the frequencies of 0.10 and 0.11  $\text{yr}^{-1}$  due to the climatically

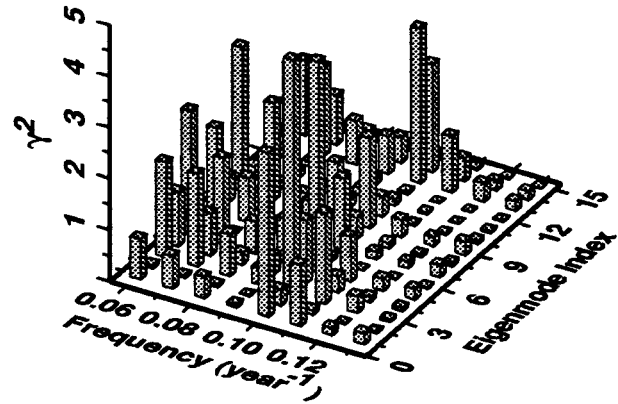


FIG. 11. Contributions to  $\gamma^2$  using all 36 detection boxes (global) for the EBM response to the combination of all four climate forcings, SVGA.

significant volcanic eruptions over the last 100 yr (Fig. 7). In addition, the greenhouse gas forcing has significant power over the entire solar band of frequencies (0.06 to 0.13  $\text{yr}^{-1}$ ). Either of these two climate forcings is much larger than the solar-cycle forcing over the solar band.

The wider implication of these findings is that all four deterministic climate forcings need to be included in any signal detection effort. If in searching for one such signal the other signals are treated as simply part of the background climate noise, serious contamination of the results may occur. Since all of the signals have some commonality in their climatic space-time patterns, a detection method must be used that isolates the signal-of-interest from the influence of the other climate signals.

Using our method to isolate the individual climate signals, we cannot conclude with much confidence that the solar-cycle signal has been detected in the observed data. This does not rule out its detection in other datasets. However, we conclude that the climate signals due to volcanic aerosols, greenhouse gases, and anthropogenic aerosols have been detected in the observed data with a very high level of confidence ( $\geq 99.97\%$ ). It is important to note that even if our estimate of the total error variance given in (40) is underestimated by a factor of 2, the confidence level for the detection of the greenhouse gas signal in the observed data is only reduced to 99.93%. Finally, we conclude that the signal due to the combination of the four climate forcings considered in this study is present in the observed data at a confidence level of 99.9%. It should be pointed out that this detection method cannot distinguish between two climate forcings, which have the same space-time response pattern. Thus, any climate forcing that is similar to the greenhouse gas forcing over the past 100 yr would be mistakenly attributed to greenhouse gas forcing.

*Acknowledgments.* This work was supported by a

grant from the National Institute for Global Environmental Change (Department of Energy) through its South Central Regional Office at Tulane University. G. R. North was also supported by a Department of Energy CHAMMP grant. The Department of Energy does not necessarily endorse the findings of this study.

## REFERENCES

- 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.
- Boucher, O., and U. Lohmann, 1995: The sulfate-CCN-cloud albedo effect. A sensitivity study with two general circulation models. *Tellus*, **47B**, 281–300.
- Hasselmann, K., 1979: On the signal-to-noise problem in atmospheric response studies. *Meteorology over the Tropical Oceans*, D. B. Shaw, Ed., Royal Meteorological Society, 251–259.
- , 1993: Optimal fingerprints for the detection of time-dependent climate change. *J. Climate*, **6**, 1957–1971.
- , 1997: Multi-pattern fingerprint method for detection and attribution of climate change. *Climate Dyn.*, **13**, 601–611.
- Haywood, J. M., R. J. Stouffer, R. T. Wetherald, S. Manabe, and V. Ramaswamy, 1997: Transient response of a coupled model to estimated changes in greenhouse gas and sulfate concentrations. *Geophys. Res. Lett.*, **24**, 1335–1338.
- Hegerl, G. C., and G. R. North, 1997: Comparison of statistically optimal approaches to detecting anthropogenic climate change. *J. Climate*, **10**, 1125–1133.
- , 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.
- , K. Hasselmann, U. Cubasch, J. F. B. Mitchell, E. Roeckner, R. Voss, and J. Waszkewitz, 1997: Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change. *Climate Dyn.*, **13**, 613–634.
- IPCC, 1996: *Climate Change 1995: The Science of Climate Change*. J. T. Houghton, L. G. Meira Filho, B. A. Callander, N. Harris, A. Kattenberg, and K. Maskell, Eds., Cambridge University Press, 584 pp.
- Jones, A., D. L. Roberts, and A. Slingo, 1994: A climate model study of indirect radiative forcing by anthropogenic sulphate aerosols. *Nature*, **370**, 450–453.
- Jones, P. D., 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.
- Kiehl, J. T., and B. P. Briegleb, 1993: The relative roles of sulfate aerosols and greenhouse gases in climate forcing. *Science*, **260**, 311–314.
- Kim, K.-Y., 1996: Sensitivity of a linear detection procedure to the accuracy of empirical orthogonal functions. *J. Geophys. Res.*, **101**, 23 423–23 432.
- , G. R. North, and J. Huang, 1996: EOFs of one-dimensional cyclostationary time series: Computations, examples, and stochastic modeling. *J. Atmos. Sci.*, **53**, 1007–1017.
- Manabe, S., and R. J. Stouffer, 1996: Low-frequency variability of surface air temperature in a 1000-yr integration of a coupled atmosphere–ocean–land surface model. *J. Climate*, **9**, 376–393.
- North, G. R., and K.-Y. Kim, 1995: Detection of forced climate signals. Part II: Simulation results. *J. Climate*, **8**, 409–417.
- , —, S. S. P. Shen, and J. W. Hardin, 1995: Detection of forced climate signals. Part I: Filter theory. *J. Climate*, **8**, 401–408.
- Ramaswamy, V., and C.-T. Chen, 1997: Linear additivity of climate response for combined albedo and greenhouse perturbations. *Geophys. Res. Lett.*, **24**, 567–570.
- Santer, B. D., K. E. Taylor, T. M. L. Wigley, J. E. Penner, P. D. Jones, and U. Cubasch, 1995: Towards the detection and attribution of an anthropogenic effect on climate. *Climate Dyn.*, **12**, 77–100.
- , and Coauthors, 1996: A search for human influences on the thermal structure of the atmosphere. *Nature*, **382**, 39–46.
- 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.
- Thiébaux, H. J., 1994: *Statistical Data Analysis for Ocean and Atmospheric Sciences*. Academic Press, 247 pp.