ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 23, NO. 1, 2006, 61–68

Statistical Procedures for Estimating and Detecting Climate Changes

S. S. P. SHEN*

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Canada

(Received 20 December 2004; revised 5 August 2005)

ABSTRACT

This paper provides a concise description of the philosophy, mathematics, and algorithms for estimating, detecting, and attributing climate changes. The estimation follows the spectral method by using empirical orthogonal functions, also called the method of reduced space optimal averaging. The detection follows the linear regression method, which can be found in most textbooks about multivariate statistical techniques. The detection algorithms are described by using the space-time approach to avoid the non-stationarity problem. The paper includes (1) the optimal averaging method for minimizing the uncertainties of the global change estimate, (2) the weighted least square detection of both single and multiple signals, (3) numerical examples, and (4) the limitations of the linear optimal averaging and detection methods.

Key words: Climate change detection, optimal averaging, optimal detection, mean square error, multivariate analysis, linear regression, inference

1. Introduction

The detection of forced climate change signals involves three tasks: (1) identifying climate changes such as global warming, (2) detecting the existence of forced climate signals in the observed climate data, and (3) attributing the climate changes to known forcings. Tasks (2) and (3) are often merged in a single mathematical procedure. The solution to the detection problem thus answers the questions of how much warming (or cooling) has happened in the recent past on both regional and global scales, and whether the 20th-century warming has been caused by anthoropogenic activities. The statistical procedures of signal detection in communication engineering are used in detecting climate changes, and the procedures combine signal filtering, climate model simulation, and climate observations. Gerald North of the United States and Klaus Hassalmann of Germany pioneered the development of the theory of optimally detecting climate change signals (North et al., 1995; Hassalmann, 1993). Their theories have different mathematical expressions, but the procedures of their methods are the same and contain the following five steps.

(1) Simulate the natural variations of climate, such as El Niño, by using climate models without forcing. The result is the background noise.

(2) Simulate the forced climate signals by using climate models with external forcing such as the man-made greenhouse gases. The result is the forced climate signal, or the fingerprint of the external forcing on the climate.

(3) Organize the observed climate data in the same way as is used to organize the simulated forced signal data.

(4) Apply a weighted linear regression to the observed data and to the signal data.

(5) Use statistical inference to decide whether the forced climate signals contained in the observed data are significant at a given significance level.

For the layman, the above procedures can be explained as follows. Ten-year-old John had his three best friends over for a Saturday-night sleepover in his room. His father slept next door to monitor the situation. John and his friends not only talked until after midnight, but also spoke loudly and, worse, played loud music. Often two or more boys talked simultaneously. From the sound next door, the father concluded that the boys were getting more and more excited. He then attempted to detect who was talking and what was being talked about. He was familiar with the music and each boy's voice, and tried to use the sounds of several voices to identify who was the main cause of the excitement. In terms of the above

*E-mail: shen@ualberta.ca

climate change detection procedures, the father's assessment of the sound next door corresponds to the optimal assessment of the rising global surface air temperature (section 2 of this paper). His familiarity with the music and other noises (i.e., the background noise) corresponds to the noise simulation step (1) above. His knowledge of the boys' voices (i.e., the signal from each boy) corresponds to the signal simulation step (2). The father's comparison of what he heard with the boys' voice characteristics corresponds to the signal representation step (3) (i.e., the representation of climate signals and noise in the same framework). The father's separation of the boys' voices from the noises corresponds to the regression detection step (4). Finally, the father's conclusion about who was the main cause of the continued excitement corresponds to the inference or attribution step (5).

Numerous papers on climate change detection have been published since the early 1990s. However, the statistical and computational procedures have not been clear to many scientists in the climate research community, and a concise summary of the statistical procedures for a general audience interested in climate research is still lacking. The purpose of this paper is to provide such a short summary of optimal averaging, detection, and attribution, so that a research scientist working on climate change detection can easily follow the procedures presented here and avoid reading unnecessary or even confusing mathematical formulas and statements. The paper pays particular attention to the assumptions of optimal averaging and linear regression modeling, and to the space-time approaches in the signal detection procedures. It also discusses the limitations of the linear detection method. However, this paper does not intend to provide a review of the literature on climate change detection, since they are documented in IPCC (2001, chapters 2 and 12). A long summary of the detection procedures is available in Zwiers (1999), and detailed computational examples can be found in North and Wu (2001).

The remainder of this paper is arranged as follows. Section 2 describes an optimal estimation of the climate-warming signal. Section 3 provides the algorithm for detecting a single signal, while section 4 deals with multiple signals. Section 5 describes a numerical example in North and Wu (2001). Section 6 provides conclusions and a discussion.

2. Optimal estimate of global warming

The purpose of this section is to identify climate changes by optimally estimating the global average surface air temperature anomalies (SATA). Doing so is the first step of detection: identifying the climate change.

The apparent warming trend of the surface air temperature (SAT) since the 1970s in most inhabited parts of the Northern Hemisphere suggested the need for a careful study of the global average annual mean of the SATA. Historically, the surface air and water temperature were observed at discrete points by fixed stations or moving vessels. Thus the true SATA global average defined by

$$\bar{T}(t) = \frac{1}{A}\int_{\Omega} T(\hat{\boldsymbol{r}}, t) d\Omega \qquad (1)$$

must be estimated by discrete data, where $T(\hat{\boldsymbol{r}}, t)$ denotes the SATA field, Ω the region under investigation, and A the area of Ω. The task is to optimally estimate the global average from the observed data to minimize the estimate's uncertainties. The IPCC (2001) used an optimal averaging (OA) method based on empirical orthogonal functions (EOFs) to calculate the global average and provided an error bound (Fig. 2.8 of IPCC 2001, chapter 2; Folland et al., 2001; Shen et al., 1994 and 1998). The global average is often calculated from the temperature anomaly data on grid boxes. Let T_i be the SATA data on the grid box i, $\langle E_i^2 \rangle$ be the data error variance for the same box, and $\langle \cdot \rangle$ be the ensemble average. The estimated average of Eq. (1) is then

$$\hat{\bar{T}} = \sum_{i \in N} w_i T_i \ , \qquad (2)$$

with a normalization condition on the weights

$$\sum_{i \in N} w_i = 1 \ ,$$

where N denotes the set of grid boxes with observed data, and w_i is the weight for the SATA data on the grid box i. The mean square error (MSE) between the true average $\overline{T}$ and the estimated average $\hat{\overline{T}}$ has an EOF representation given by

$$\varepsilon^2 = \langle (\overline{T} - \hat{\overline{T}})^2 \rangle \approx \sum_{n=1}^{M} \lambda_n (\bar{\psi}_n - \hat{\bar{\psi}}_n)^2 + \sum_{i \in N} w_i^2 \langle E_i^2 \rangle \ , \qquad (3)$$

where M is the number of EOFs used, λ_n is the eigenvalue of the n EOF mode, $\psi_n(i)$ is the value of the nth EOF mode in box $i, \overline{\psi}_n$ is the "true" spatial average of the nth order EOF mode and is computed by

$$\overline{\psi}_n = \sum_{j \in G} A_j \psi_n(j) \ , \qquad (4)$$

and $\hat{\overline{\psi}}_n$ is the estimated spatial average of EOF mode n given by

$$\hat{\overline{\psi}}_n = \sum_{i \in N} w_i \psi_n(i) \ . \qquad (5)$$

Here, G denotes the entire grid, and A_j is the area of the grid box j ($j \in G$) (Shen et al., 1998).

The error formula (3) has interesting interpretations. First, the error is decomposed into two parts. The first part is the sampling error due to the presence of climate teleconnections expressed in terms of the EOF patterns, and the second is the random error of uncorrelated observational noise consisting of observational errors. A special case is the sampling of a spatially-white noise field with a constant observational noise. In this case, the optimal weights are $1/N$, and the MSE becomes the familiar sampling error formula MSE=σ^2/N, where σ^2 is the error variance of the observed data.

Second, the term in the square brackets is the numerical integration error of the EOF patterns. The sampling error of the average temperature is thus converted into the sampling error of the EOFs and the uncorrelated observational noise.

Third, the first part of the sampling error formula indicates that the sampling error is linearly proportional to the eigenvalues λ_n and is in a quadratic relationship with the sampling error of the EOF patterns. This universal physical phenomenon is associated with spectral representation (i.e., eigenvalue and eigenfunction expansion): the accuracy of the eigenvalues is more important than the eigenfunctions in estimating errors. When considering the integrated behavior of an electron and an atom in, for example, radiation, an electron's energy levels are certainly more important than its exact orbit. Thus, exploring the various methods that can yield the accurate eigenvalues of a climate system is worthwhile. An extrapolation method to refine the eigenvalues of sea surface temperature has been tested and has proved effective (Shen et al., 1998; Lin, 2002)

The minimization condition of MSE leads to a group of linear equations that determine the optimal weights w_i ($i = 1, 2, \ldots, N$) and the Lagrange multiplier Λ in the constrained optimization:

$$\sum_{j \in N} \rho_{ij} w_j + \langle E_i^2 \rangle w_i + \Lambda = \bar{\rho}_i \ , \quad i = 1, 2, \ldots, N, \tag{6}$$

$$\sum_{j \in N} w_j = 1 \ , \tag{7}$$

where

$$\rho_{ij} \approx \sum_{n=1}^{M} \hat{\lambda}_n \psi_n(i) \psi_n(j) \ . \tag{8}$$

is the approximate covariance matrix, and

$$\bar{\rho}_i \approx \sum_{n=1}^{M} \hat{\lambda}_n \psi_n(i) \bar{\psi}_n \tag{9}$$

is the approximate spatial average of the covariance function over the grid box i. Here, $\hat{\lambda}_n$ is the approximate value of λ_n. These eigenvalues and eigenfunctions may vary from year to year. Folland et al. (2001) thus used a moving time window approach to calculate $\hat{\lambda}_n$ and $\psi_n(i)$ for every year.

Computationally, these equations may become ill-conditioned when N is too large and M is too small. The observational error variance $\langle E_i^2 \rangle$ helps to eliminate the ill condition, particularly when these error variances are large. Folland et al. (2001) used the annual $5^\circ \times 5^\circ$ latitude-longitude grid-box HadCRUTv anomalies (relative to the 1961–1990 average) and the corresponding error variances to create averaged anomalies and error variances on a $10^\circ \times 20^\circ$ latitude-longitude grid. The number N of data boxes was thus reduced, and the resulting set of linear equa-

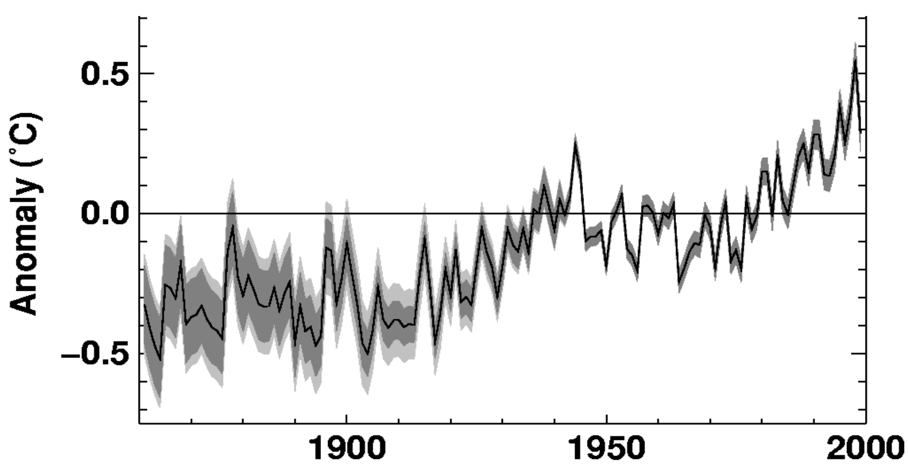

Fig. 1. Annual global SATA relative to 1961–2000 with 2σ-wide confidence intervals. Uncertainties are shown including (light shading) and excluding (darker shading) those due to changes in thermometer exposures. The overall change from a linear regression of the optimally averaged SATA from 1861 to 2000 is 0.61°C±0.16°C.

tions (6) and (7) has a unique solution. The global average annual mean of SATA and its MSE according formulas (2) and (3) are shown in Fig. 1 [Fig. 3b of Folland et al. (2001) and Fig. 2.8 of IPCC (2001)].

The above summarizes the optimal estimate of the history of the global average and hence gives a signal of the global climate change. The next question is whether this change is the result of external forcings, such as an increase of carbon dioxide in the air.

3. Detecting and attributing a single signal

The signal detection of climate change is to find whether a given set of observed climate data contains a climate signal forced by a specific external forcing such as doubled carbon dioxide concentration. The observed climate random variable T and the simulated signal random variable S are sampled, respectively, at the N spatial sites in the detection region and Y years in the detection period. The essential data are thus the $K = N \times Y$ pairs of the observed SATA $T(j,k)$ values and the simulated forced SATA signal $S(j,k)$ values, $j = 1,\ldots,N, k = 1,\ldots,Y$ defined on the (j,k) space-time grid denoted by H. The detection is set up as a regression model between the observed SATA and the simulated SATA:

$$T = \alpha S + E\,. \tag{10}$$

Due to the strong spatial inhomogeneity of surface air temperature anomalies, the residue random variable E most likely carries heteroskedasticity in this regression model, hence it does not have a uniform variance at all the data points on the grid H, and the E's at different space-time grid points may not even be independent from each other. Thus, the assumptions of the linear regression model are violated. Consequently, the linear regression results for the inference about the slope α are not valid. However, a weighted regression in the spectral space has a better chance to satisfy the assumptions of the linear regression model. Graybill and Iyer's (1994, section 3.3, section 8.2) book is a good reference for the assumptions of linear regression and the weighted least square regression.

The data are projected onto the modes of the background noise simulated by control runs of climate models. The mth eigenvalue and eigenvector of the space-time covariance matrix $\boldsymbol{\Sigma}_{K\times K}$ of the background noise are λ_m and $\boldsymbol{B}_m$, respectively, where $\boldsymbol{B}_m$ is a $K \times 1$ vector defined on the space-time grid H. The projected data, i.e., the Fourier coefficients based on the modes $\boldsymbol{B}_m$, are used as the data, which are T_m and $S_m (m = 1, 2, \ldots, M)$. Here, M is the truncation order of the eigenvector expansion. The strength of the signal simulated by a climate model is measured by the "theoretical" signal-to-noise ratio γ with

$$\gamma^2 = \sum_{m=1}^{M} \frac{S_m^2}{\lambda_m}\,.$$

This ratio depends solely on the accuracy of the climate model and is independent of the observed climate data. When $\gamma > 2.0$, we usually say that the climate model can simulate the signal. However, γ^2 is an increasing function of the mode truncation order M. In order to correctly use this formula to calculate the signal-to-noise ratio, we should also justify that our calculation is robust. The justification can be done by calculating γ^2 with different truncation orders.

Let $E_m = T_m - \alpha S_m$. Transform the spectral data in the regression form into

$$\frac{T_m}{\sqrt{\lambda_m}} = \alpha \frac{S_m}{\sqrt{\lambda_m}} + \frac{E_m}{\sqrt{\lambda_m}}\,.$$

Assume that the transformed residue vector is distributed as $[E_m/\sqrt{\lambda_m}] \sim N(0, \xi^2 \boldsymbol{I})$, where $\boldsymbol{I}$ is the identity matrix of order M, and $\xi^2 = \langle E_m^2 \rangle / \lambda_m$ is the ratio of the variance of the residue's mth mode to the variance of the background noise's mth mode. This ratio is usually regarded as unity. The linear regression model (10) is valid for the M pairs of transformed spectral data $(T_m/\sqrt{\lambda_m}, S_m/\sqrt{\lambda_m}), m = 1,\ldots,M$. The weighted least square approach can now be applied to minimize $\sum_m (E_m/\sqrt{\lambda_m})^2$; i.e.,

$$\min \sum_m \frac{1}{\lambda_m}(T_m - \alpha S_m)^2\,.$$

This implies

$$\hat{\alpha} = \frac{\sum_{m=1}^{M} T_m S_m/\lambda_m}{\gamma^2}\,. \tag{11}$$

If $\xi^2 = 1$, the confidence interval of the signal amplitude α at the $(1-p)100\%$ confidence level is

$$[\hat{\alpha} - z_{1-(p/2)}(1/\gamma), \alpha + z_{1-(p/2)}(1/\gamma)]\,, \tag{12}$$

where $z_{1-(p/2)}$ is the upper $100(p/2)$ percentile value of the standard normal distribution. If ξ^2 is not unity and has to be estimated, then the confidence interval is

$$[\hat{\alpha} - t_{1-(p/2),M-2}\mathrm{SE}(\alpha), \alpha + t_{1-(p/2),M-2}\mathrm{SE}(\alpha)]\,,$$

where

$$\mathrm{SE}(\alpha) = \sqrt{\frac{\sum_{m=1}^{M} \frac{1}{\lambda_m}(T_m - \alpha S_m)^2}{(M-2)\gamma^2}}$$

is the standard error of the slope, and $t_{1-(p/2),M-2}$ is the upper $100(p/2)$ percentile value of the t-distribution with $M-2$ degrees of freedom.

The slope $\hat{\alpha}$ is the signal amplitude contained in the observed data of T. If the confidence interval of $\hat{\alpha}$

does not contain zero, then the signal amplitude is significantly different from zero at the given significance level p100%, and one can conclude that the observed data contain the climate change signal due to a specific forcing, say, greenhouse gases.

4. Detecting multiple signals

Climate change signals are rarely isolated, for they occur together. The most important forcing signals are the changes due to G (greenhouse gases), V (volcanic eruptions), S (solar radiation variation), and A (aerosol particles in the atmosphere). The corresponding signals are denoted by $S_{\rm G}, S_{\rm V}, S_{\rm S}$ and $S_{\rm A}$, respectively. The background noise field is assumed to be the same as that described in the last section. The signal data in the spectral space are $S_{m,\rm G}, S_{m,\rm V}, S_{m,\rm S}$ and $S_{m,\rm A}(m = 1, 2, \ldots, M)$. The observational data are the same as those in the last section $T_m(m = 1, 2, \ldots, M)$. For the same reasons as in the last section, to satisfy the assumptions of linear regression, the multivariate linear regression model is assumed in the spectral space and is estimated by using the weighted least square method. The model is

$$T = \sum_n b_n S_n + E \ , \tag{13}$$

where n=G, V, S, A.

We assume $[E_m/\sqrt{\lambda_m}] \sim N(\mathbf{0}, \xi^2\boldsymbol{I})$, just as we did for a single signal in the previous section. The signal-to-noise ratio γ_k is still determined by

$$\gamma_n^2 = \sum_{m=1}^{M} \frac{(S_{m,n})^2}{\lambda_m} \ , \quad n = \text{G, V, S, A} \ .$$

The weighted least squares

$$\min \sum_m \frac{1}{\lambda_m} \left(T_m - \sum_n b_n S_{m,n} \right)^2 \ ,$$

leads to the estimates of the slope vector

$$\hat{\boldsymbol{b}} = (\boldsymbol{S}'\boldsymbol{W}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}\boldsymbol{T}_{\rm D} \ . \tag{14}$$

Here, the prime "′" denotes the matrix transpose, the matrix

$$\boldsymbol{S} = \begin{pmatrix} S_{1,\rm G} & S_{1,\rm V} & S_{1,\rm S} & S_{1,\rm A} \\ S_{2,\rm G} & S_{1,\rm V} & S_{2,\rm S} & S_{2,\rm A} \\ & \cdots & \cdots & \\ S_{M,\rm G} & S_{M,\rm V} & S_{M,\rm S} & S_{M,\rm A} \end{pmatrix}$$

is the signal data, the vector $\hat{\boldsymbol{b}}' = (\hat{b}_{\rm G}\ \hat{b}_{\rm V}\ \hat{b}_{\rm S}\ \hat{b}_{\rm A})$ is the estimator of $\boldsymbol{b}' = (b_{\rm G}\ b_{\rm V}\ b_{\rm S}\ b_{\rm A})$, the vector $\boldsymbol{T}'_{\rm D} = (T_1 T_2 \ldots T_M)$ is the observational data, and the diagonal matrix

$$\boldsymbol{W} = [W_{ij}]_{M\times M} = [\delta_{ij}/\lambda_i]$$

is the weight characterizing the noise level. In the above, δ_{ij} is the Kronecker delta, which is equal to one if $i = j$, and zero otherwise.

For the estimate $\hat{\boldsymbol{b}}$, a confidence hyper-ellipsoid can be calculated with a given confidence level. If the ellipsoid does not intersect with the coordinate hyperplane, the signal in the component orthogonal to the plane is detected from the observational data.

In the case of four signals as discussed above and of $\xi^2 = 1$, the $100(1-p)$% confidence ellipsoid is determined by the following formula:

$$(\boldsymbol{b} - \hat{\boldsymbol{b}})'\boldsymbol{S}'\boldsymbol{W}\boldsymbol{S}(\boldsymbol{b} - \hat{\boldsymbol{b}}) \leqslant \chi_4^2(p) \ , \tag{15}$$

where $\chi_4^2(p)$ is the upper $(100p)$th percentile of a χ^2-distribution with 4 degrees of freedom [see equation (A12.1.4) of IPCC (2001, p.732)]. If ξ^2 is not unity and is to be estimated, then

$$(\boldsymbol{b} - \hat{\boldsymbol{b}})'\boldsymbol{S}'\boldsymbol{W}\boldsymbol{S}(\boldsymbol{b} - \hat{\boldsymbol{b}}) \leqslant 4s^2 F_{4,M-4}(p) \ , \tag{16}$$

where $F_{4,M-4}(p)$ is the upper $(100p)$th percentile of an F-distribution with $(4, M-4)$ degrees of freedom, and

$$s^2 = (\boldsymbol{T}_{\rm D} - \boldsymbol{S}\hat{\boldsymbol{b}})'\boldsymbol{W}(\boldsymbol{T}_{\rm D} - \boldsymbol{S}\hat{\boldsymbol{b}})/(M-4)$$

is an estimate of ξ^2. When the truncation order M is very large, say, 30, then $4s^2F_{4,M-4}(p) \approx \chi_4^2(p)$, where p is usually between 0.01 and 0.1. The confidence ellipsoid (16) is approximated by that of (15).

Johnson and Wichern's (1992, section 7.4) book is a good reference for the above results. The above statistical formulas have been widely used in the existing detection literature, but in the present summary, both the mathematics and interpretation of the detection have been simplified and made transparent and simple for applications.

5. Numerical procedures and examples

Here, we present a simple and straightforward numerical procedure. Computations of this detection procedure are made on grid boxes. Three datasets are needed:

(1) Observational data on selected grid boxes in the detection period.

(2) Signal data on the same boxes and in the same period from model simulations with specific forcings.

(3) Noise data on the same boxes and in the same period from controlled model runs.

The computation steps are as follows:

(1) Construct a space-time state vector. Suppose six grid boxes and fifty years of annual mean surface air temperature anomalies are chosen as the climate state variables. The space-time state vector is then a 300 × 1, where 300 = 6 × 50. $\boldsymbol{U}' = [U_1^1 U_2^1 \ldots U_{50}^1 \ldots\ldots U_1^6 U_2^6 \ldots U_{50}^6]$.

(2) The data of many years of controlled runs of climate models, say 10,000 years, are divided into 200, 50-year sections. Identify the data of the controlled

runs for the space-time state vector and construct a covariance matrix for the vector. The matrix's order is 300×300, and the rank is at most 200 (because of the 200 sections of the controlled runs and because $200 < 300$). The eigenvalues (λ_m) and eigenvectors ($\boldsymbol{B}_m$) of the matrix can be computed. Only the first $M(\leqslant 200)$ eigenvalues and eigenvectors are retained.

(3) Organize the signal data and observed data according to the space-time state vector U and compute the principal components (i.e., the dot product between the signal data and observational data with the EOFs $\boldsymbol{B}_m$). This procedure gives S_m and T_m $(m = 1, \ldots, M)$.

(4) Compute the signal-to-noise ratio γ_n and see if the climate model can generate the signal due to a specific forcing.

(5) Compute the regression coefficients $\hat{\boldsymbol{b}}$ according to Eq.(14).

(6) Make an inference about the signal in the observational data according to the confidence ellipsoid or confidence interval [see Eqs.(15) or (16)]. The conclusion is that either a signal is detected in the observational data with a certain significance level, or that the signal is not detected.

The following numerical examples are adopted from North and Wu (2001). Seventy-two $10° \times 10°$ grid boxes are chosen: Box 1 (15°S, 10°W), Box 2 (20°N, 20°W), ..., Box 36 (40°S, 175°E), where the latitude and longitude are the coordinates of the center of the grid box. The observational annual mean surface air temperature anomaly data on the grid boxes are from the Jones $5° \times 5°$ dataset. The $10° \times 10°$ data are derived by averaging the data over four neighboring $5° \times 5°$ grid boxes. The controlled runs are from several GCM models, each running for 1,000 years. The combination of the runs is used as the background noise data. North and Wu (2001) considered signals simulated by various climate models. A group of signals generated by the Hadley Center's Climate Model 2 (HadCM2) is taken as an example here. The signal to be detected occurred in the period of 1947–1996 (50 years). The signal-to-noise ratio and the regression coefficients and their confidence intervals (at the 90% confidence level) are shown in Table 1.

According to this table, the HadCM2 model shows strong greenhouse gas and volcano signals, and these signals are being successfully detected from the observational data. The model shows a moderately strong aerosol signal, but this signal has not been detected from the observational data. The model shows a very weak solar signal in the 50-year period, and this weak signal is not detected.

Table 1. Summary of the detection results for the forced climate signals sampled in the 72 grid boxes for the period of 1944–1993 by using the HadCM2 model.

Signal	γ_n	b_n	Confidence Interval
G	5.53	0.72	(0.62, 0.82)
V	3.65	0.67	(0.38, 0.96)
S	0.77	1.79	(−0.12, 2.76)
A	2.32	0.12	(−0.14, 0.38)

In the case that γ_n is very small (less than 1.0) and the confidence interval of $\hat{b}_n$ does not include zero, we will still conclude that the signal is not detected from the observational data, for we cannot confidently say that the climate model itself generates a valid signal.

6. Conclusions and discussion

This paper has described an optimal averaging method to estimate climate changes and a linear detection technique to detect specific climate change signals contained in observational data. The signal is simulated by climate models with specific external forcings. The background noise is simulated by controlled runs of climate models. The theoretical signal-to-noise ratio independent of the data shows the strength of a specific signal manifested by a climate model. A weighted least square regression is applied in the spectral space with the EOF basis and yields an estimate of the amplitudes of specific signals in the observational data. For a given confidence level, if the confidence interval for a regression coefficient does not include zero, then the corresponding specific signal is said to have been detected from the data with a certain significance level.

While detecting forced climate signals in the spectral space has advantages, such as dimension reduction, carrying out the optimal detection in the space-time physical space is more straightforward, particularly when the number of spatial sampling points N is not very large. One can transform the linear regression models (10) and (13) in the physical space to remove the heteroskedasticity and then carry out a weighted least squares estimation. Let $\boldsymbol{T}, \boldsymbol{R}$, and $\boldsymbol{E}$ be the K-dimensional vectors of observed data, signal data, and regression residue, respectively, on the space-time grid H. The background noise's space-time covariance matrix $\boldsymbol{\Sigma}_{K\times K}$ is used to transform $\boldsymbol{E}$ so that $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{E} \sim N(0, \varsigma^2\boldsymbol{I})$, where ς^2 is the ratio of the residue variance to the background noise variance and is usually regarded as unity, and $\boldsymbol{I}$ is a Kth order identity matrix. The weighted least square algorithm minimizes $\boldsymbol{E}'\boldsymbol{\Sigma}^{-1}\boldsymbol{E}$; i.e.,

$$\min(\boldsymbol{T} - \alpha\boldsymbol{R})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{T} - \alpha\boldsymbol{R}) . \tag{17}$$

This leads to an estimate of the signal amplitude

$$\hat{\alpha} = \frac{\boldsymbol{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{R}}{r_{\rm sn}} , \tag{18}$$

where r_{sn} is the signal-to-noise ratio defined in the space-time physical space:

$$r_{\text{sn}} = \boldsymbol{R}'\boldsymbol{\Sigma}^{-1}\boldsymbol{R} \,. \tag{19}$$

If ς^2 is unity, then the confidence interval for the amplitude of the single signal is

$$|\alpha - \hat{\alpha}| < z_{1-(p/2)}(1/\sqrt{r_{\text{sn}}}) \,. \tag{20}$$

If ς^2 is not unity and has to be estimated, then

$$|\alpha - \hat{\alpha}| < t_{1-(p/2),K-1}\sqrt{\frac{(\boldsymbol{T}-\alpha\boldsymbol{R})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{T}-\alpha\boldsymbol{R})}{(K-1)r_{\text{sn}}}} \,. \tag{21}$$

Similarly, one can derive the approximate formula for the amplitude of four signals. The signal amplitude vector is

$$\hat{\boldsymbol{b}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{T} \,, \tag{22}$$

where $\boldsymbol{X}$ is the $K \times 4$ signal data matrix.

If ς^2 is unity, the confidence ellipsoid for the signal amplitude $\boldsymbol{b}$ is determined by

$$(\boldsymbol{b}-\hat{\boldsymbol{b}})'(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})(\boldsymbol{b}-\hat{\boldsymbol{b}}) \leqslant \chi^2_4(p) \,. \tag{23}$$

If ς^2 is not unity and has to be estimated, then the confidence ellipsoid for $\boldsymbol{b}$ is determined by an F-distribution:

$$(\boldsymbol{b}-\hat{\boldsymbol{b}})'(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})(\boldsymbol{b}-\hat{\boldsymbol{b}}) \leqslant 4(\boldsymbol{T}-\boldsymbol{X}'\hat{\boldsymbol{b}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{T}-\boldsymbol{X}'\hat{\boldsymbol{b}})F_{4,K-4}(p) \,. \tag{24}$$

One can prove that the signal amplitude and the signal-to-noise ratio calculated in the physical space are the same as those calculated from the spectral space in sections 3 and 4 when ignoring the truncation errors in the eigenvector expansion.

Numerous other problems remain in climate change detection and attribution. The first lies in the errors contained in the three essential datasets used in our detection procedures. Errors in climate models (model errors), errors in computing the EOF coefficients T_m from the station-based observational data (sampling errors), and an insufficient length of the controlled simulation for computing the covariance matrix of the noise field (the EOF error) can contribute to the uncertainty of a conclusion. The fidelity of climate models and the accuracy of the observational data are of crucial importance in correct detection. The accuracy of regional climate models still needs to be improved; hence, the detection of regional climate change has more uncertainties than detection on the global scale.

The assumptions of the linear regression model under the transform in either the spectral space or the physical space need to be systematically checked. The check is to validate the linearity, uniform variance, and independence of the residues. There are standard statistical procedures to perform the model checking (Johnson and Wichern, 1992, pp. 308–314). In fact, the linear superposition of the climate signals expressed in the regression model (13) is problematic since climate signals have strong nonlinear interactions because the climate models are built on the first principles of physics and the parameterization of chemistry and physiology. However, in some circumstances, the linear model (13) can be a good approximation for some climate models when the forcing amplitude is small (Gillet et al., 2004).

Thus, after learning from signal analysis in communication engineering, we should develop a detection method that does not depend on climate models as crucially as the linear regression approach described here. The nonlinearity and non-stationarity of the climate process in the data must be reflected in the new method. Experimentation in detection-without-a-model has been conducted by Oh et al. (2003) using wavelet analysis and statistical regression. The recently-developed Hilbert-Huang Transform (HHT) may be another effective tool for the development of detection-without-a-model and for the consideration of both nonlinearity and non-stationarity (Huang et al., 2003).

Acknowledgments. This work was partially supported by the Meteorological Service of Canada, the NOAA Office of Global Programs, and the Chinese Academy of Sciences. This research was initiated in 2001 from discussions with Francis Zwiers and Xuebin Zhang of the Meteorological Service of Canada. Discussions with Peter Hooper of the University of Alberta on regression and inference were very helpful in the improvement of this paper.

REFERENCES

Folland, C. K., and Coauthors, 2001: Global temperature change and its uncertainties since 1861. *Geophys. Res. Lett.*, **28**, 2621–2624.

Gillett, N. P., M. F. Wehner, S. F. B. Tett, and A. J. Weaver, 2004: Testing the linearity of the response to combined greenhouse gas and sulfate aerosol forcing. *Geophys. Res. Lett.*, **31**(14), L14201, doi:10.1029/2004GL020111.

Grayhill, F. A., and H. K. Iyer, 1994: *Regression Analysis: Concepts and Applications*. Duxbury Press, California, 701pp.

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

Huang, N. E., M. C. Wu, S. R. Long, S. S. P. Shen, N. H. Hsu, D. Xiong, W. Qu, and P. Gloersen, 2003: On the establishment of a confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proc. Roy. Soc. London. Series A.*, **459**, 2317–2345.

IPCC, 2001: *Climate Change 2001: The Scientific Basis*. Houghton et al., Eds., Cambridge University Press, Cambridge, United kingdom and New York, USA, 881pp.

Johnson, R. A., and D. W. Wichern, 1992: *Applied Multivariate Statistical Analysis*. Prentice Hall, New Jersey, 642pp.

Lin, Q., 2002: High performance finite element methods. *Recent Progress in Computational and Applied PDES*. Proc. of the International Symposium on computational and Applied PDES, July 2001, Zhangjiajie, China, Kluwer Academic Publishers, New York, 267–286.

North, G. R. , K. Y. Kim, S. S. P. Shen and J. W. Hardin, 1995: Detection of forced climate signals. Part I: Theory. *J. Climate*, **8**, 401–408.

North, G. R., and Q. Wu, 2001: Detecting climate signals using space-time EOFs. *J. Climate*, **14**, 121–145.

Oh, H. S., C. Ammann, P. Naveau, D. Nychka, and B. Otto-Bliesner, 2003: Multi-resolution time series analysis applied to solar irradiance and climate reconstructions. *J. Atmospheric and Solar-Terrestrial Phys.*, **65**, 191–201.

Shen, S. S. P., G. R. North, and K. Y. Kim, 1994: Spectral approach to optimal estimation of the global average temperature. *J. Climate*, **7**, 1999–2007.

Shen, S. S. P., T. M. Smith, C. F. Ropelewski, and R.E. Livezey, 1998: An optimal regional averaging method with error estimates and a test using tropical Pacific SST data. *J. Climate*, **11**, 2340–2350.

Zwiers, F. W., 1999: Climate Change Detection: A review of techniques and applications. 1999: Anthropogenic Climate Change, Proc. of the First GKSS Spring School on Environmental Research, von Storch et al., Eds., Springer Verlag, 161–203.