Detection of Forced Climate Signals. Part I: Filter Theory

GERALD R. NORTH

Climate System Research Program, College of Geosciences and Maritime Studies, Texas A&M University, College Station, Texas

KWANG-Y. KIM*

Applied Research Corporation, College Station, Texas

SAMUEL S. P. SHEN

Department of Mathematics, University of Alberta, Edmonton, Canada

JAMES W. HARDIN

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

(Manuscript received 21 December 1992, in final form 29 April 1994)

ABSTRACT

This paper considers the construction of a linear smoothing filter for estimation of the forced part of a change in a climatological field such as the surface temperature. The filter is optimal in the sense that it suppresses the natural variability or "noise" relative to the forced part or "signal" to the maximum extent possible. The technique is adapted from standard signal processing theory. The present treatment takes into account the spatial as well as the temporal variability of both the signal and the noise. In this paper we take the signal's waveform in space–time to be a given deterministic field in space and time. Formulation of the expression for the minimum mean-squared error for the problem together with a no-bias constraint leads to an integral equation whose solution is the filter. The problem can be solved analytically in terms of the space–time empirical orthogonal function basis set and its eigenvalue spectrum for the natural fluctuations and the projection amplitudes of the signal onto these eigenfunctions. The optimal filter does not depend on the strength of the assumed waveform used in its construction. A lesser mean-square error in estimating the signal occurs when the space–time spectral characteristics of the signal and the noise are highly dissimilar; for example, if the signal is concentrated in a very narrow spectral band and the noise in a very broad band. A few pedagogical exercises suggest that these techniques might be useful in practical situations.

1. Introduction

The problem of detecting a climate change signal in the climatological record is of obvious importance in any strategy to understand global change. Considering its importance, relatively few attempts have been made to define the problem rigorously. Several approaches have been developed but so far the problem has not yielded to a completely satisfactory analysis.

Hasselmann (1979) was among the first to discuss the problem of investigating forced climate responses with signal processing methods. He suggested representing the forced signal in terms of the EOF (empirical orthogonal functions) basis set. Data can be projected onto the EOF components for two different times in history. The square of the difference of these amplitudes divided by the corresponding eigenvalues is a positive definite random variable. Taking the null hypothesis to be that no signal is present in the data stream, this random variable is approximately a χ^2 variate. A χ^2 test can in principle be used to test the null hypothesis. The EOF basis plays the same role in space as the Fourier frequency basis does along the time axis for a stationary process. Hasselmann (1979, 1993) suggested a "signal processing" technique in which a prescription is made for filtering out the EOF components in the data stream that have no signal projection on them. Filtering out the uninteresting EOF components reduces the number of degrees of freedom in the χ^2 test. By such a prefiltering scheme we can enhance the portion of the data stream we are looking for while systematically discarding the parts in which we are uninterested. A similar type of procedure will come out naturally in the present paper.

* Current affiliation: Climate System Research Program, Texas A&M University, College Station, Texas.

Corresponding author address: Prof. Gerald R. North, College of Geosciences, Texas A&M University, College Station, TX 77843-3150.

In a very interesting approach, Bell (1982, 1986) constructed a single optimally weighted scalar index based on a weighted sum over data that provided a maximum signal-to-noise ratio. He tailored the weighting scheme according to information based upon model predictions. Bell's numerical examples illustrated that considerable enhancement could be gained in the early detection of global warming by optimally weighting the observations. The optimal weighting procedure used by Bell to "detect" a forced climate change is similar to the philosophy used in the present paper.

Solow (1991) presented a conventional multivariate statistical procedure. He considers difference vectors between the control run of a model and data. For example, one component of the vector might correspond to each station where data are collected. Let the difference vector of dimension N be given by $\mathbf{Y} - \mathbf{a}$. The variate

$$X = (\mathbf{Y} - \mathbf{a})'\mathcal{C}^{-1}(\mathbf{Y} - \mathbf{a}),$$

where $\mathcal{C}$ is the covariance matrix of the natural variability between the various stations. The EOF basis set is the one that renders $\mathcal{C}$ diagonal. Therefore,

$$X = \sum_{n=1}^{N} \frac{T_n^2}{\lambda_n},$$

where T_n is the projection of the data onto the nth EOF, and λ_n is the corresponding eigenvalue. Under the null hypothesis that no climate change occurs, the variate X will be distributed as χ_N^2.

While having the virtue of objectivity, the χ^2 method makes little use of the prior knowledge possessed by the researcher about the nature of the expected signal or of the natural variability to obtain an optimal estimate of the signal strength. A counterintuitive element arises when we try to decide on how to choose the summation cutoff N. In principle (in the absence of signal processing), N should be infinite, since we are considering a continuous random field (e.g., surface temperature). Climatologists know that variance (and information content) in the large index EOF components is usually very small ($\lambda_n \rightarrow 0$ as $n \rightarrow \infty$). But the χ^2 formalism as stated above treats the last component in the sum exactly as the first. This suggests that a different formalism might be called for in the climate change problem.

Barnett (1986, 1991), Barnett and Schlesinger (1987), and Barnett et al. (1991) have introduced and examined a "fingerprinting" technique. They advocated construction of a pattern correlation index. This index is a measure of the correlation between the observed data stream and a predicted signal. A high pattern correlation would indicate confirmation of the hypothesis that the model-predicted signal is being observed. In these studies some mention is made of the prefiltering of the data stream onto the EOFs as a means of enhancing the signal-to-noise ratio (hereafter we use the shorthand SNR to denote signal-to-noise ratio), but since the work was preliminary, explicit use of signal processing has not yet been reported. The efforts so far with the fingerprinting technique have been unsuccessful in detecting a greenhouse signal. The reasons could be several (aside from the trivial one that the greenhouse effect does not exist). 1) The signal processing technique was not adequate; that is, the SNR was simply too small for the sample. 2) Except in Barnett (1991), the hypothesis was tested against equilibrium runs and the potential signatures are quite different in equilibrium versus transient runs. 3) Strong contributors to natural variability in the data are the ENSO phenomena, which are in virtually no models used in greenhouse simulations. 4) The strongest feature in the data record of globally averaged temperatures is a cooling in the period mid-1940s to mid-1970s and no greenhouse model to date is able to simulate such a dip. (A cursory look at the data suggests that the dip is dominated by events in the Arctic, possibly sea ice related, and current models are unable to simulate such a fluctuation with any reliability.)

The present paper attempts to develop a signal processing prescription that can be used along with some prior knowledge about the predicted signal to enhance the signal in the data relative to the natural variability. This is accomplished by constructing a filter such that when applied to the data stream yields an estimate of the signal that is best in the least mean square error (MSE) sense. The problem is solved by taking the variation of the MSE, setting it to zero, and solving the resulting integral equation. The EOF basis set in space and time is found automatically to be the natural basis set to use in the solution of this problem. The filter selectively passes EOF components where the signal is large compared to the noise, and it weighs less those where this is less so. This is accomplished in such way as to not bias the estimate. (We postpone the study of biased estimators that might in principle lower the MSE even further.) To find such a filter, of course, we have to know something about the signal beforehand. The trial signal may be given as deterministic from a model simulation or it may be allowed to have some uncertainty allowing for our lack of its knowledge (at the price of a higher MSE).

Our experience with climate models suggests that the surface temperature field is one of the most robust fields that can be simulated in climate change experiments (see e.g., North et al. 1982; North et al. 1983a; Kim and North 1991; Leung and North 1991; North et al. 1992). Hence, it might be useful to look for characteristic patterns in the surface temperature as a trial candidate for a forced climate signal. The steady increase of the globally averaged temperature comes to mind first. This is indeed the single largest signal we are likely to see in the response, since increases in CO_2 force the global-scale temperature field directly. Almost all theories also predict polar amplification of the thermal response due to one form or another of the ice feedback mechanisms. Although large, the polar amplification is problematic because the spatial pattern of the response is very model dependent, since current models do not handle sea-ice processes reliably.

There is evidence to suggest that the response of the temperature field to a time-dependent large spatial-scale forcing is fairly robust in the absence of ice feedback mechanisms (e.g., Kim et al. 1992). The main feature is that the land temperatures tend to lead the ocean surface temperatures in phase and have larger amplitude. The seasonal cycle is a trivial example, but the phenomenon persists even at very low frequencies (e.g., Kim et al. 1992). Such presumably robust spatial patterns correlated with frequency might provide a basis for a signal processing scheme.

Our goal in this series of papers is to see if signal processing, which exploits both the temporal and spatial waveforms, can be applied in such a way as to improve our ability to detect forced signals embedded in natural variability. We note that the techniques of optimal filtering are very general and apply to a much broader range of problems than those introduced here merely as examples. We also admit that ours is only a minor early step, since many more sophisticated techniques have evolved in the last half-century in the field of signal processing that atmospheric scientists are just beginning to appreciate. While the present paper presents some examples of primarily pedagogical value, Part II of this sequence (North and Kim 1995; hereafter referred to as II) will present numerical simulations based on the energy balance models mentioned above.

2. Defining a filter

Consider a model of the surface temperature field,

$$T(\hat{\mathbf{r}}, t) = T_S(\hat{\mathbf{r}}, t) + T_N(\hat{\mathbf{r}}, t), \qquad (1)$$

where we have made a separation of the field into two parts: a forced part or "signal" $T_S(\hat{\mathbf{r}}, t)$ and a natural variability or "noise" part $T_N(\hat{\mathbf{r}}, t)$. Implicit is the assumption that the two are independent (at least for small T_S). This is in some sense an assumption about the linearity of the system. We, of course, know that this is not likely to hold for large signals, but some GCM modeling evidence suggests that at least for the atmospheric part of the system the separation for surface temperature is valid (North et al. 1992). For the cases of interest, $T_S(\hat{\mathbf{r}}, t)$ is to be composed of mainly low frequencies (periods of order decadal) and large scales ($\gg$1000 km), while $T_N(\hat{\mathbf{r}}, t)$ has significant variance at all frequencies (red noise spectrum) with autocorrelation time for large-scale land-covered areas of the order of a month. The spatial decomposition of $T_N(\hat{\mathbf{r}}, t)$ is essentially white for low spherical harmonic indices and turning over at spherical harmonic degree approximately 8 (Leung and North 1991; North et al. 1992). We use spherical harmonics here only as a guide, since $T_N(\hat{\mathbf{r}}, t)$ should be decomposed into a set of space–time empirical orthogonal functions (stEOFs). Such a stEOF decomposition has been carried out for the surface temperature field and will be used as an example in the present sequence of papers (Kim and North 1993).

We seek a smoothing filter $\Gamma(\hat{\mathbf{r}}, t; \hat{\mathbf{r}}', t')$ to apply to our data to estimate the amplitude of $T_S(\hat{\mathbf{r}}, t)$ in the least-mean-square-error sense. We denote our estimator by $\hat{T}_S(\hat{\mathbf{r}}, t)$:

$$\hat{T}_S(\hat{\mathbf{r}}, t) = \int_{\mathcal{T}} \int_{\mathcal{D}} \Gamma(\hat{\mathbf{r}}, t; \hat{\mathbf{r}}', t') T(\hat{\mathbf{r}}', t') d\Omega' dt', \qquad (2)$$

where $\mathcal{D}$ is some domain on the earth (possibly the whole earth) and the temporal integral runs over an appropriate interval $\mathcal{T}$, for example, some segment of the immediate past. In this formulation we have assumed that the field has been measured continuously in both space and time. In practice, the effects of multiple point or splotchy data collection designs can be taken into account as will be shown in later sections.

Hence, the problem as posed is to find a good smoothing filter. This problem is well known in the field of signal processing for the time domain (e.g., Franks 1981; Wiener 1949; Kolmogorov 1941) or in image processing for the space domain (e.g., Jain 1989). In the engineering terminology we wish to determine as well as possible the strength of a deterministic continuous waveform in space and time. We acknowledge that there are model examples of dynamical systems in which the ideal separation of the signal from the noise we speak of here will not hold.

Experience with the complicated notation involved suggests that from the outset we set $(\hat{\mathbf{r}}, t) \rightarrow x$ and use the simplifications $T_S(\hat{\mathbf{r}}, t) \rightarrow T_S$, $T_S(\hat{\mathbf{r}}', t') \rightarrow T_S(x') \rightarrow T'_S$. Similarly, $d\Omega dt \rightarrow dx$, etc. In this notation (2) becomes

$$\hat{T}_S = \iint \Gamma(x, x') \{ T'_S + T'_N \} dx'. \qquad (3)$$

3. Mean square error (MSE) and constraint

To obtain a measure of the performance of a particular smoothing filter consider the MSE:

$$\epsilon^2 = \langle (T_S - \hat{T}_S)^2 \rangle, \qquad (4)$$

where $\langle \cdot \rangle$ denotes ensemble averaging. How can we adjust the functional form of $\Gamma(x, x')$ to make ϵ^2 least? If $\Gamma(x, x')$ were given as a trial form with only a few free parameters, we would have $\epsilon^2(\alpha_1, \alpha_2, \ldots, \alpha_n)$, where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the adjustable parameters. Then we could set the partial derivatives of ϵ^2 separately to zero and solve the resulting algebraic equations for the parameters. There is usually an additional algebraic constraint condition on the parameters to ensure that the estimate is unbiased.

a. No-bias constraint

If we take the ensemble average of (3), we should get the true value of T_S:

$$\langle \hat{T}_S \rangle = T_S, \tag{5}$$

or

$$T_S = \int \Gamma(x, x') T'_S dx'. \tag{6}$$

This means that our final optimal form for $\Gamma(x, x')$ will have to be such that Eq. (6) is satisfied. In the present formulation we will take $T_S(x)$ to be a deterministic signal computable by a model. In principle we could allow $T_S(x)$ also to be stochastic but for now we take it to be deterministic and its space–time shape to be given. Finally, note that Eq. (6) appears to be an equation determining $\Gamma(x, x')$. The obvious solution would be $\delta(x - x')$, the unit operator. This is, of course, useless since the filter would accomplish no purification whatsoever. The fact is Eq. (6) is not to be satisfied for any function $T_S(x)$ but rather will depend explicitly upon $T_S(x)$. The filter will have to be tailored to satisfy Eq. (6), even if it leads to an explicit dependence on the signal we are trying to detect.

b. Minimizing the MSE

Now we insert the definitions into Eq. (4) and find

$$\epsilon^2 = T_S^2 + \iint \Gamma(x, x'')\Gamma(x, x') \times \{ T''_S T'_S + \langle T''_N T'_N \rangle \} dx' dx'' - 2T_S \int \Gamma(x, x') T'_S dx'.$$

To set up the minimization problem subject to the no-bias constraint, it is convenient to use the method of Lagrange multipliers (Arfken 1985). To facilitate this we introduce a functional of the filter $J[\Gamma]$, defined by

$$J[\Gamma] = \epsilon^2 - 2\Lambda\left\{ T_S - \int \Gamma(x, x') T'_S dx' \right\},$$

where 2Λ is a Lagrange multiplier.

In finding the optimal estimator we must set

$$\delta J = 0.$$

We consider the variable in the variation to be $\Gamma(x, x')$. Then

$$\delta J = 2 \iint \delta\Gamma(x, x')\Gamma(x, x') \times \{ T''_S T'_S + \langle T''_N T'_S \rangle \} dx dx' - 2T_S \int \delta\Gamma(x, x') T'_S dx' + 2\Lambda \int \delta\Gamma(x, x') T'_S dx' = 0. \tag{7}$$

The last can hold only for arbitrary $\delta\Gamma$ if the coefficient of it is identically zero. This leads to

$$T_S T'_S - \Lambda T'_S = \int \Gamma(x, x'') \{ T''_S T'_S + \langle T''_N T'_N \rangle \} dx'', \tag{8}$$

which is an integral equation to be solved for the optimal $\Gamma(x, x')$. Note the peculiar dependence on the left-hand argument x; it is not integrated over. Hence, we can think of it as a label. To emphasize this, write it as $\Gamma_x(x')$, and in some cases the x dependence will be suppressed altogether in the notation. Then after a cancellation, Eq. (8) reduces to

$$\int_{\mathcal{D}\times T} \rho(x'', x') \Gamma_x(x'') dx'' = -\Lambda T'_S, \tag{9}$$

where

$$\rho(x', x'') = \langle T_N(x') T_N(x'') \rangle \tag{10}$$

is the space–time covariance of the natural variability. The integral equation for the optimal filter is similar to the relation found by Bell (1986) in his optimal weighting scheme.

4. Solving for the optimal filter

Equation (9) constitutes an integral equation for $\Gamma_x(x')$, the desired smoothing filter. To solve the integral equation, we introduce a basis set of functions that allows inversion of the equation to find an explicit formula for the filter.

a. Space–time EOFs

The stEOFs are the eigenfunctions of the covariance kernel:

$$\int_{\mathcal{D}\times T} \rho(x', x'') \psi_n(x'') dx'' = \lambda_n \psi_n(x'). \tag{11}$$

They are orthonormal,

$$\int_{\mathcal{D}\times T} \psi_n(x') \psi_m(x') dx' = \delta_{nm}, \tag{12}$$

and complete

$$\sum_n \psi_n(x') \psi_n(x'') = \delta(x' - x''). \tag{13}$$

The covariance can be expressed

$$\rho(x', x'') = \sum_n \lambda_n \psi_n(x') \psi_n(x''). \tag{14}$$

The reader is cautioned that the index n above consists of both space and time eigenvalue indices. For example, if the noise process is stationary in time (and if the time integral runs from $-\infty$ to ∞), the Fourier integral representation is the appropriate temporal part of the stEOF basis (see Wallace and Dickinson 1972; North

1984; Hasselmann 1987). Then such quantities as (Σ_n) imply an integral over frequencies or in practice a sum over discrete frequency bands. For each frequency band there will be a countable infinity of discrete stEOF indices for the spatial part of the problem. Hence, although the notation is very compact, it might mask a considerable computational effort. We emphasize that if the time integral runs only over a finite temporal segment $\mathcal{T}$, the Fourier integral basis is not appropriate, although it often might be a useful approximation.

b. stEOF basis for $\Gamma_x(x')$

We can expand our unknown

$$\Gamma(x') = \sum_n \Gamma_n \psi_n(x'),$$

where the x dependence of $\Gamma_x(x') \equiv \Gamma'$ and $\Gamma_n(x)$ has been suppressed.

c. Computation of Γ, Λ

Using the expansion of $\Gamma(x')$ in the expression for the least error Eq. (9), we have

$$\Lambda T_{Sm} = -\lambda_m \Gamma_m,$$

or

$$\Gamma_m = -\Lambda \frac{T_{Sm}}{\lambda_m}.$$

Now we decompose Γ'

$$\Gamma' = \sum_m \Gamma_m \psi'_m = -\Lambda \sum_m \frac{T_{Sm}\psi'_m}{\lambda_m}.$$

To evaluate the Lagrange parameter Λ we must satisfy the constraint Eq. (6). After insertion we find

$$\Lambda = -\frac{T_S}{\gamma^2},$$

where

$$\gamma^2 = \sum_n \frac{T_{Sn}^2}{\lambda_n}. \qquad (15)$$

Note that the Lagrange parameter Λ is a function of x, since the original signal T_S is.

We may now write the smoothing filter explicitly:

$$\Gamma(x, x') = \frac{T_S(x)}{\gamma^2} \sum_m \frac{T_{Sm}\psi_m(x')}{\lambda_m}. \qquad (16)$$

The filter at space–time location x is proportional to the shape of the predicted signal shape $T_S(x)$; the proportionality factor is related to the projection of the signal onto the stEOF basis and weighted inversely with the noise variance attributed to that mode in the sum. The estimator is then

$$\hat{T}_S(x) = \frac{T_S(x)}{\gamma^2} \sum_m \frac{T_{Sm}}{\lambda_m} T_m, \qquad (17)$$

where T_m is the projection amplitude of the data onto the mth stEOF. Since $\langle T_m \rangle = T_{Sm}$, it is obvious that the estimator is unbiased. Note that the overall scale of the signal function cancels out, since it occurs squared in the numerator and also squared in the quantity γ^2.

We note with interest the similarity of this estimator to the "fingerprint" method introduced by Barnett (1986). The similarity lies in the numerator of the individual terms in the sum. We see products like $T_{Sm}T_m$ normalized by λ_m. These terms are each proportional to the pattern covariance for that stEOF component. In our case, however, larger λ_m will tend to suppress the contribution to a given model.

d. Expression for MSE

It is useful at this point to compute the actual optimal MSE. We have after some manipulations and cancellations

$$\epsilon^2 = \iint \Gamma(x, x')\Gamma(x, x'')\rho(x', x'')dx'dx'' = \sum_n \lambda_n \Gamma_n^2.$$

After insertion and further manipulations we obtain the very simple form

$$\epsilon^2 = \frac{T_S^2}{\gamma^2}, \qquad (18)$$

where γ^2 is given by Eq. (15). Clearly the quantity γ^2 will control the size of the MSE in the problem. For good resolution of the signal we prefer $\gamma \gg 1$.

e. Signal to noise

By rearranging and taking the square root we see that the parameter γ is the theoretical signal-to-noise ratio, which has the same value at each point in space and time. We want to make γ as large as possible. Referring to Eq. (15), we see that this can be accomplished by having the T_{Sm}^2 large for values of m for which λ_m is small. Note that these sequences are constrained by their sums being fixed:

$$\sum_m \lambda_m = \text{total variance of } T_N = \int \langle T_N(x)^2 \rangle dx, \qquad (19)$$

and

$$\sum_m T_{Sm}^2 = \text{total variance of } T_S = \int T_S^2(x)dx. \qquad (20)$$

The ideal situation for signal detection is for the signal to occupy only a few stEOF modes and for the eigenvalue spectrum of the noise to be very small in those signal modes. Then when we filter the whole data stream so that only those modes are retained, we will be assured of having a larger SNR. In addition to filtering out unwanted parts of the spectral components of the data stream, the filtered parts are weighted so as to emphasize the parts where the signal is largest.

5. Illustrative examples

a. Deterministic sharp tone

Sunspots pose an interesting example of a weak-tone forcing to climate. Recent satellite data suggest that the 11-year cycle of sunspots is correlated with an oscillation of the solar constant whose amplitude is about 0.1% (see Willson and Hudson 1991). This leads to a variation of the global average temperature of the order of 0.02°C (North et al. 1983b). This is a very faint signal to look for considering that the standard deviation of the natural variability of annual-averaged data is about nine times this (excluding El Niño) (Kim and North 1992). No signal processing leads then to a signal-to-noise ratio of 0.11, which is far short of being satisfactory. In this pedagogical signal processing example, we pretend that the signal is a pure tone and that we have 10 cycles of data. Spectral lines in 100 years of data are 0.01 yr^{-1} apart. Just to have some numbers to work with we can take our natural variability to be red noise turning over at a frequency of $\pm(5\ \mathrm{yr})^{-1}$. If we divide the natural variability evenly over the 41 lines in this low-frequency band and examine the signal (0.02°C) contribution added to the line at $f_s = (10\ \mathrm{yr})^{-1}$, we obtain an SNR of 0.35, three times the unfiltered value.

A further enhancement can be found by taking into account the deterministic phase of the response. In the periodogram, a single spectral line is actually composed of two parts, the sine and cosine parts. Suppose the phase is known. Then we can take the signal's phase to be pure sine and project out the cosine part with our filter. This can be used to remove half the noise variance. Hence, the phase information can lead to a further factor of $\sqrt{2}$ enhancement in the SNR. Therefore, an ideal filter of the globally averaged temperature response might conceivably be detected at a level of SNR of about 0.5 with a century of data. Even though we have amplified the SNR a factor of 5, we still have very weak performance. Further enhancement can be obtained through the use of the geographical distribution of the signal at the forcing frequency as will be demonstrated in II. Of course, in the real case the tone is not at a discrete frequency but rather it is clustered in the neighborhood of $f_s \sim (11\ \mathrm{yr})^{-1}$. A further refinement to the filter involves inclusion of seasonal dependence of the signal and the noise (signal stronger in summer, noise weaker). The seasonal effect can be accommodated using the formalism developed in this paper and will be the subject of future studies.

Naturally other factors will enter besides the raw SNR parameter. For example, should we use a broader-band filter to reduce sampling variability? As we use a slightly broader-band filter, we will weaken the detected signal but we simultaneously reduce the uncertainty in the noise for a single realization. These are typical tradeoffs faced by the investigator in this type of signal processing exercise. We leave these issues for future studies special to the sunspot problem.

b. Spatially sharp signal

As a purely pedagogical example, we consider a very symmetrical problem with a particular time-independent spatial pattern forcing. First omit the time dependence. We take the spatial variability to be homogeneous on the sphere. This means that the EOFs are the spherical harmonics. In this example, suppose we want to detect a signal consisting of a single spherical harmonic component $T_S(\hat{\mathbf{r}}) = \alpha Y_{33}(\hat{\mathbf{r}})$. The quantity γ^2 can be computed:

$$\gamma^2 = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{T_{Sm}}{\lambda_{nm}} = \frac{\alpha^2}{\lambda_{33}}.$$

Now α is the projection of the signal on the noise stEOF basis and λ_{33} is the fraction of the total variance spectrum explained by the $Y_{33}(\hat{\mathbf{r}})$ mode. If the noise (natural variability) is very broad band (spread over many stEOF modes), then λ_{33} will be small. In this case the SNR resulting from use of the optimal filter will be $1/\lambda_{33}$. Hence, great performance can be expected from the optimal filter when the noise is very broad band and the signal is very narrow band. This principle will apply in time as well as space. The sunspot problem in 5a is an example of a narrowband signal, for which the optimal filter greatly improves the SNR. In time, a narrow concentration of variance in the signal will imply a quasi-periodic process.

Next consider the optimal filter itself for the example just raised. It is

$$\Gamma(\hat{\mathbf{r}}, \hat{\mathbf{r}}') = \frac{T_S(\hat{\mathbf{r}})}{\gamma^2} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{T_{S_{nm}}}{\lambda_{nm}} Y_{nm}(\hat{\mathbf{r}}')$$
$$= Y_{33}(\hat{\mathbf{r}}) Y_{33}(\hat{\mathbf{r}}').$$

Hence, the filter takes the whole measured field $T_S(\hat{\mathbf{r}}) + T_N(\hat{\mathbf{r}})$ and projects it onto the (3, 3) spherical harmonic mode. This leads to precisely the correct shape (after all we forced that) but gives an erroneous amplitude. Consider a particular realization in which the amplitude of $Y_{33}(\hat{\mathbf{r}})$ in the noise field is β times a random number with ensemble-mean zero and variance λ_{33}. Then for this particular realization, the estimator of $T_S(\hat{\mathbf{r}})$ is

$$\hat{T}_S(\hat{\mathbf{r}}) = (\alpha + \beta) Y_{33}(\hat{\mathbf{r}}).$$

This is clearly an unbiased estimator since $\langle \beta \rangle = 0$. The MSE is $\lambda_{33}|Y_{33}|^2$ as predicted. We note that as expected, Λ is independent of the signal strength.

Unfortunately, the sharp signal case (signal composed of only one spectral component) is so degenerate that the optimality aspects are trivial. Hence, we turn to an example with a two-mode signal.

c. A two mode signal

We take the signal to be $T_S(x) = \alpha_1\psi_1(x) + \alpha_2\psi_2(x)$. Then the SNR squared γ^2 is given by

$$\gamma^2 = \frac{\alpha_1^2}{\lambda_1} + \frac{\alpha_2^2}{\lambda_2}.$$

The total variance of the signal is $\alpha_1^2 + \alpha_2^2$, while the noise variance is $\sum_{n=0}^{\infty} \lambda_n$, of which λ_1 and λ_2 are only two terms. This suggests that γ^2 is much larger than the ratio of the overall signal variance to the overall noise variance. In this case, the MSE at point x is

$$\epsilon^2 = \frac{(\alpha_1\psi_1(x) + \alpha_2\psi_2(x))^2}{\dfrac{\alpha_1^2}{\lambda_1} + \dfrac{\alpha_2^2}{\lambda_2}}.$$

The optimal smoothing filter is given by

$$\Gamma(x, x') = \frac{T_S(x)}{\gamma^2}\left\{\frac{\alpha_1}{\lambda_1}\psi_1(x') + \frac{\alpha_2}{\lambda_2}\psi_2(x')\right\}.$$

If we assume that for this particular realization $T_N(x) = \beta_1\psi_1(x) + \beta_2\psi_2(x)$, where the β_i are random numbers with mean zero and variance λ_i, then we can write the expression for our estimator

$$\hat{T}_S(x) = \frac{T_S(x)}{\gamma^2}\left\{\frac{\alpha_1}{\lambda_1}(\alpha_1 + \beta_1) + \frac{\alpha_2}{\lambda_2}(\alpha_2 + \beta_2)\right\},$$

which is easily seen to be unbiased (γ^2 equals the quantity in brackets if the $\beta_{1,2}$ vanish). This example shows that the weighting coefficients of the stEOF signal amplitudes are such as to weight according to the expected SNR in contrast to the trivial sharp-signal case. As before, we note that α_1 and α_2 are each proportional to the a priori assumed overall signal strength or scale. The squared dependence in the numerator will be precisely cancelled by that in the parameter γ^2.

6. Summary and conclusions

In this paper we considered the problem of a stream of data in space and time in which there is a deterministic climatic signal embedded in a natural variability field. Our model of the process is that the two be additive and uncorrelated. Our purpose was to design a filter that when applied to the stream provided a minimum MSE estimator of the signal. We developed the formalism for an unbiased estimator. The problem was cast into the form of an integral equation for the filter that could be solved by use of the stEOF functions for the natural variability.

The general finding is that such filters are relatively easy to construct, if the stEOFs are known in advance. Such stEOFs can be computed from long runs of coupled ocean–atmosphere GCMs or for pilot studies in simpler models (Kim and North 1991, 1992, 1993). We provided some simple pedagogical examples of filters: 1) a pure tone imposed on a red noise background (numbers relevant to the sunspot forcing of the climate system were used), 2) a pure spatial (spherical harmonic) tone on a noisy sphere, and 3) a combination tone on the noisy sphere.

It would be remiss of us to not point out the many problems encountered in applying our formalism to real climate change problems. For example, the real observing system has many defects (biases and sampling gaps) that will decrease the SNR. However, we believe that model studies in conjunction with space–time signal processing methodology can assist in deciding what datasets to gather and how to treat them in the detection problem.

Part II of this sequence (following paper in this issue of the *Journal of Climate*) will consider the signal processing problem in some examples of forced signals in stochastic energy balance model simulations. In these examples we can get a clearer idea of how the space–time patterns of forced signals can be exploited to improve the SNR.

Acknowledgments. We thank the Department of Energy for its support of this work via a grant to Applied Research Corporation through CHAMMP (DE-FG05-91ER61221). One of us (GRN) has also benefited from a contract from Battelle Pacific Northwest Laboratories. Neither the Department of Energy nor Battelle necessarily endorse any of the conclusions drawn in the paper. We also wish to thank Thomas Crowley and William Pennell for many helpful discussions. We also benefited greatly from the referees' comments.

REFERENCES

Arfken, G., 1985: *Mathematical Methods for Physicists*. Academic Press, 985 pp.

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

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

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

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

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.

Franks, L. E., 1981: *Signal Theory.* Dowden and Culver, 317 pp.

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.

——, 1987: PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.,* **93,** 11 015–11 021.

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

Jain, A. K., 1989: *Fundamentals of Image Processing.* Prentice Hall, 569 pp.

Kim, K.-Y., and G. R. North, 1991: Surface temperature fluctuations in a stochastic climate model. *J. Geophys. Res.,* **96,** 18 573–18 580.

——, and ——, 1992: Seasonal cycle and second-moment statistics of a simple coupled climate system. *J. Geophys. Res.,* **97,** 20 437–20 448.

——, and ——, 1993: EOF analysis of surface temperature field in a stochastic climate model. *J. Climate,* **6,** 1681–1690.

——, ——, and J. Huang, 1992: On the transient response of a simple coupled climate system. *J. Geophys. Res.,* **97,** 10 069–10 081.

Kolmogorov, A., 1941: Interpolation and extrapolation of stationary time series. *Bulletin de L'Academie des Sciences de U.S.S.R.,* Ser. Math. 5, 3–14. (in Russian).

Leung, L., and G. R. North, 1991: Atmospheric variability on a zonally symmetric all-land planet. *J. Climate,* **4,** 753–765.

North, G. R., 1984: Empirical orthogonal functions and normal modes. *J. Atmos. Sci.,* **41,** 879–887.

——, and K. Y. Kim, 1995: Detection of forced climate signals. Part II: Simulation results. *J. Climate,* **8,** 409–417.

——, T. L. Bell, R. F. Cahalan, and F. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.,* **110,** 699–706.

——, J. Mengel, and D. Short, 1983a: Simple energy balance model resolving the seasons and the continents: Application to the astronomical theory of the ice ages. *J. Geophys. Res.,* **88,** 6576–6586.

——, ——, and ——, 1983b: Climatic response to a time varying solar constant. *Weather and Climate Responses to Solar Variations.* B. M. McCormac, Ed., Colorado Associated University Press, 243–255.

——, K.-Y. Yip, R. Leung, and R. Chervin, 1992: Forced and free variations of the surface temperature field in a GCM. *J. Climate,* **5,** 227–239.

Solow, A., 1991: On the statistical comparison of climate model output and climate data. *Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations.* M. E. Schlesinger, Ed., Elsevier, 505–510.

Wallace, J. M., and R. Dickinson, 1972: Empirical orthogonal representation of time series in the frequency domain. Part I: Theoretical considerations. *J. Appl. Meteor.,* **11,** 887–892.

Wiener, N., 1949: *Time Series.* The MIT Press, 163 pp.

Willson, R. C., and H. S. Hudson, 1991: The sun's luminosity over a complete solar cycle. *Nature,* **351,** 42–44.