# Predictive Oscillation Patterns: A Synthesis of Methods for Spatial-Temporal Decomposition of Random Fields

Charles KOOPERBERG and Finbarr O'SULLIVAN

Spatial-temporal decompositions of climatologic fields have been obtained using a range of techniques, including principal component analysis (PCA) and principal oscillation patterns (POPS). PCA decompositions are forced to be correlated to the original field, but they may not capture interesting aspects of temporal variation. On the other hand, POPS decompositions focus on temporal variation but are not forced to correlate to the field. Here we introduce a hybrid of these methods that attempts to retain desirable aspects of both PCA and POPS. The approach attempts to project the field onto a lower dimensional subspace with the property that the average error associated with forecasting a future state of the field on the basis of the history contained in the projection is minimized. A recursive algorithm for estimating a spatial-temporal decomposition based on this idea is developed. The methodology is applied to a 47-year climatological record of the 5-day average 500-millibar-height anomaly field, sampled on a 445 grid over the Northern Hemisphere extra-tropics. Some asymptotic properties of the estimation method for the new technique are examined in a simple situation. Although the estimation method requires a consistent estimator of a certain spectral density matrix, the target parameters are estimated at a parametric rate. Interestingly, the details of the nonparametric estimation of the spectral density, such as the choice of the smoothing kernel, do not appear to affect the asymptotic variance of the target parameters.

KEY WORDS: Geopotential height anomaly; One-step ahead forecast error; Principal components; Principal oscillation patterns; Spectral density matrix.

## 1. INTRODUCTION

The study of fluid motion is of substantial interest in atmospheric science and oceanography (Chelton 1994; Panel on Statistics and Oceanography 1994). Although physical principles describing the small-scale or short-term evolution of the fluid are relatively well established (Holton 1992; Wallace and Hobbs 1977), summary characterization of the low-frequency behavior in terms of spatial and temporal variation are less well understood. As a result, there has been ongoing interest in empirical methods capable of summarizing phenomenological aspects of fluid motion. One thrust has been to separate spatial and temporal variation via representations consisting of a limited number of fixed spatial patterns, each pattern with its own characteristic temporal forcing. Such representations have been achieved using a collection of statistical of techniques, including principal components analysis (PCA) (Barnett and Preisendorfer 1987; Wallace, Smith, and Bretherton 1992). A more recent innovation in the atmospheric science literature is a technique known as principal oscillation patterns analysis (POPS) (Hasselmann 1988; Von Storch 1994; Von Storch, Bruns, Fischer-Bruns, and Hasselmann 1988). This method incorporates first order Markov modeling assumptions to isolate spatial patterns with a strong temporal dependence. The original motivation for our work was to develop an alternative to POPS that would be less reliant on the first-order Markov assumption.

Let $\{x(t); t = \ldots, -1, 0, 1, 2, \ldots,\}$ be a discrete mean zero vector-valued real time series of dimension $p$ representing the field under study. In an atmospheric science setting, the components of $x$ correspond to spatial locations where the field is observed; see Section 3 for an example. It is assumed that the measurements have been corrected for nonstationary (temporal) trends and, perhaps by complex demodulation, band-pass filtered to a range of temporal scales of interest (Bretherton, Smith, and Wallace 1991; Von Storch et al. 1988). The general form of the type of spatial-temporal representations considered in the atmospheric science literature and in this article is

$$x(t) = \mathbf{A}z(t) + \eta^K(t) = \sum_{j=1}^{K} \mathbf{A}_j z_j(t) + \eta^K(t),$$

where $\mathbf{A}_j$ is a vector of dimension $p$ capturing a characteristic *spatial scale of variation* and $z_j(t)$ is the associated forcing function describing the temporal evolution of the field at that spatial scale. The term $\eta^K(t)$ is the residual error in the representation. Typically, $K \ll p$, so the representation provides a valuable statistical summary of the macroscopic features of the field.

In PCA the spatial patterns are derived from the spectral decomposition of the marginal spatial covariance of the field, $\mathbf{\Sigma}_0 = \text{var } x(t)$. The temporal forcing $z(t)$ are obtained by a least squares regression of $x(t)$ on the spatial patterns. Although the first $K$ principal components are defined by the property that they define the closest $K$-dimensional subspace to the $p$-dimensional marginal distribution of the field (Mardia, Kent, and Bibby 1979), there is no explicit constraint on the temporal structure of the as-

Charles Kooperberg is Assistant Professor and Finbarr O'Sullivan is Professor, Department of Statistics, University of Washington, Seattle, WA 98195. Kooperberg was supported in part by National Science Foundation grant DMS-9403371. O'Sullivan was supported in part by National Institutes of Health grant CA-57903. The authors are indebted to John M. Wallace of the Department of Atmospheric Sciences, University of Washington, for introducing them to the subject matter of this paper and for providing the geopotential height data, and to an associate editor and a referee for comments that greatly improved this article.


Journal of the American Statistical Association
December 1996, Vol. 91, No. 436, Theory and Methods

sociated forcing functions, the $z(t)$, because this definition is in terms of the marginal distribution. This could be unsatisfactory, because it means that the spatial patterns need not reflect anything to do with the evolution of the field, which is often of prime interest for modeling purposes. The frequency domain application of principal components analysis developed by Brillinger (1981) has the potential to capture interesting temporal scales; however, because this leads to separate spatial components corresponding to *each* temporal frequency, some potential for understanding the macroscopic behavior in terms of a simple parsimonious representation is lost.

A real square matrix $\mathbf{B}$ of dimension $p$ can be written as

$$\mathbf{B} = \sum_{j=1}^{p} \frac{\lambda_j}{d_j} \mathbf{R}_j \mathbf{L}_j^*,$$

where $\mathbf{R}_j$ and $\mathbf{L}_j$ are the right and left eigenvectors corresponding to the $j$th eigenvalue $\lambda_j$ (which could be complex) and $d_j = \mathbf{L}_j^* \mathbf{R}_j$, with the $^*$ indicating the complex conjugate (see Rao 1973). The matrix $\sum_{j=1}^{p} (1/d_j)\mathbf{R}_j\mathbf{L}_j^*$ is an expression for the identity matrix, and so $x(t)$ can be written as

$$x(t) = \left\{ \sum_{j=1}^{p} \frac{1}{d_j} \mathbf{R}_j \mathbf{L}_j^* \right\} x(t) = \sum_{j=1}^{p} \mathbf{R}_j \mathbf{u}_j(t),$$

with $\mathbf{u}_j(t) = \mathbf{L}_j^* x(t)/d_j$.

The POPS analysis is motivated by a first-order Markov model, AR(1), for the field (Priestley 1987, chap. 9). If the field follows the AR(1) model, then

$$x(t) = \mathbf{B}x(t-1) + \boldsymbol{\varepsilon}(t),$$

where $\boldsymbol{\varepsilon}(t)$ is assumed to be a mean zero uncorrelated (in time) process. With the foregoing decomposition of $\mathbf{B}, \mathbf{u}_j(t)$ would now satisfy a one-dimensional complex model AR(1) with coefficient given by $\lambda_j$,

$$\mathbf{u}_j(t) = \lambda_j \mathbf{u}_j(t-1) + \mathbf{L}_j^* \boldsymbol{\varepsilon}(t)/d_j.$$

The POPS analysis is an expansion for the field in terms of the right eigenvectors of an estimated AR(1) coefficient matrix, $\hat{\mathbf{B}} = \boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_0^{-1}$, where $\boldsymbol{\Sigma}_1$ is the lag-one autocovariance matrix. The eigenvectors corresponding to the largest (in modulus) $K$ eigenvalues are used. The spatial patterns are given by the vectors $\mathbf{R}_j$, and the temporal forcing functions are given by $\mathbf{L}_j^* x(t)/d_j$:

$$x(t) = \sum_{j=1}^{p} \mathbf{R}_j \mathbf{u}_j(t) + \eta^K(t) \equiv \sum_{j=1}^{K} \mathbf{A}_j z_j(t) + \eta^K(t).$$

Note that the spatial patterns and the forcing functions may be complex. Although the POPS analysis focuses on components with a strong temporal autocorrelation pattern, there is no constraint that the forcing functions be highly correlated to the original field. This complements the PCA analysis in which the correlation between $z(t)$ and $x(t)$ is maximized, but there is no guarantee that $z(t)$ has an interesting temporal structure.

Alternative analysis of lag-one autocovariance or autocorrelation matrices in terms of singular value decompositions and canonical correlations have also been proposed (see Barnett and Preisendorfer 1987, Bretherton et al. 1991, and Wallace et al. 1992). For simplicity, we limit the discussion to only PCA and POPS. The goal is to introduce a representation in which the temporal dependence characteristic of POPS and the high correlation pattern of PCA are both explicitly considered. The approach is motivated by considering the forecast error associated with using the information contained in the forcing function to predict future values of the field. The representation and estimation methodology is developed in Sections 2 and 3. An illustration with real data is given in Section 4. The dataset is a 47-year National Meteorological Center (NMC) record of 5-day average 500 hPa geopotential height field in the Northern Hemisphere extra-tropics from 1946 onward. A discussion indicating directions for future development concludes the article. Some theoretical properties of the estimation methods are provided in an Appendix.

## 2. PREDICTIVE OSCILLATION PATTERNS (PROPS)

### 2.1 Motivation: An Upper Bound on the Forecast Error

We first examine the error associated with the least squares prediction of the future state of the field, based on the information contained in the evolution of the field (at a particular spatial scale) up to the present time. Let $\mathbf{A}$ be an arbitrary $p$-dimensional unit vector describing a spatial scale of interest. The one-step-ahead forecast error associated with an optimal (in a least squares sense) forecast for the field, based on the information contained in the scale defined by $\mathbf{A}$, can be developed as follows: let $e(t+1) = x(t+1) - \mathbf{A}z(t+1) = [I - \mathbf{A}\mathbf{A}']x(t+1)$, where $z(t+1) = \mathbf{A}'x(t+1)$. Then

$$\begin{aligned} x(t+1) &- E[x(t+1)|\mathbf{H}_t(\mathbf{A})] \\ &= e(t+1) + \mathbf{A}z(t+1) \\ &\quad - E[e(t+1) + \mathbf{A}z(t+1)|\mathbf{H}_t(\mathbf{A})] \\ &= e(t+1) - E[e(t+1)|\mathbf{H}_t(\mathbf{A})] \\ &\quad + \mathbf{A}\{z(t+1) - E[z(t+1)|\mathbf{H}_t(\mathbf{A})]\}, \end{aligned}$$

where $\mathbf{H}_t(\mathbf{A}) = \{z(s) = \mathbf{A}'x(s)\colon s \le t\}$ is the history of a process up to time $t$. Because $[I - \mathbf{A}\mathbf{A}']\mathbf{A} = 0$, the two components of the forecast error are orthogonal, and so the squared length of the forecast error is given by

$$\begin{aligned} &\|x(t+1) - E[x(t+1)|\mathbf{H}_t(\mathbf{A})]\|^2 \\ &\quad = \|e(t+1) - E[e(t+1)|\mathbf{H}_t(\mathbf{A})]\|^2 \\ &\qquad + \{z(t+1) - E[z(t+1)|\mathbf{H}_t(\mathbf{A})]\}^2. \end{aligned}$$

Thus the expected forecast error $\mathrm{FE}(\mathbf{A}|x)$ is the sum of two terms, which can be reduced as follows:

$$\begin{aligned} &\mathrm{FE}(\mathbf{A}|x) \\ &\quad = E\|x(t+1) - E[x(t+1)|\mathbf{H}_t(\mathbf{A})]\|^2 \end{aligned}$$

$$
\begin{aligned}
&= E\|e(t+1) - E[e(t+1)|\mathbf{H}_t(\mathbf{A})]\|^2 \\
&\quad + E\{z(t+1) - E[z(t+1)|\mathbf{H}_t(\mathbf{A})]\}^2 \\
&= E\|e(t+1)\|^2 - E\|[e(t+1)|\mathbf{H}_t(\mathbf{A})]\|^2 \\
&\quad + E\{z(t+1) - E[z(t+1)|\mathbf{H}_t(\mathbf{A})]\}^2 \\
&= \text{trace}\{\boldsymbol{\Sigma}_0\} - \mathbf{A}'\boldsymbol{\Sigma}_0\mathbf{A} - E\|E[e(t+1)|\mathbf{H}_t(\mathbf{A})]\|^2 \\
&\quad + 2\pi \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log \mathbf{g}_{zz}(\omega)\, d\omega\right\},
\end{aligned}
$$

where $\boldsymbol{\Sigma}_0$ is the marginal covariance of the field and $\mathbf{g}_{zz}(\omega)$ is the spectral density of $z(t)$. The somewhat complicated formula for the one-step-ahead forecast error of $z(t)$ involving the spectral density has been given by Priestley (1987, chap. 10) and follows under some regularity conditions, integrability of the logarithm of the spectral density being the main condition.

Because of the way in which the history is restricted to a projection of the process, we know of no simple expression for the conditional expectation function $E[e(t+1)|\mathbf{H}_t(\mathbf{A})]$. Things do not appear to simplify even in the case when one restricts to a finite number of lag terms in the set $\mathbf{H}_t(\mathbf{A})$. Thus a workable form for the $E\|E[e(t+1)|\mathbf{H}_t(\mathbf{A})]\|^2$ term as a function of $\mathbf{A}$ is not available. In view of this, we instead consider an approximate upper bound based on dropping the complicating term. The motivation is that because $e(t+1)$ is algebraically orthogonal to $\mathbf{A}$, it seems reasonable that by a suitable choice of $\mathbf{A}$, one could make $e(t+1)$ small, and then the information about $e(t+1)$ contained in the history $\mathbf{H}_t(\mathbf{A})$ might be negligible. Thus we obtain

$$
\begin{aligned}
\text{FE}^+(\mathbf{A}|x) &= \text{trace}\{\boldsymbol{\Sigma}_0\} - \mathbf{A}'\boldsymbol{\Sigma}_0\mathbf{A} \\
&\quad + 2\pi \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log \mathbf{g}_{zz}(\omega)\, d\omega\right\} \\
&\geq E\|x(t+1) - E[x(t+1)|\mathbf{H}_t(\mathbf{A})]\|^2 \\
&= \text{FE}(\mathbf{A}|x).
\end{aligned}
$$

In Section 2.3 we show that for a simplified example, this upper bound is sharp; that is $\text{FE}^+ = \text{FE}$. Better approximations to $\text{FE}(\mathbf{A}|x)$ would be worth future consideration.

Let $\mathbf{f}_{xx}(\omega)$ denote the spectral density matrix of $x, \mathbf{g}_{zz}(\omega) = \mathbf{A}'\mathbf{f}_{xx}(\omega)\mathbf{A}$, and $\int_{-\pi}^{\pi} \mathbf{f}_{xx}(\omega)\, d\omega = \boldsymbol{\Sigma}_0$. The expression for the upper bound on the forecast error can be manipulated and reduced to the form

$$
\begin{aligned}
\text{FE}^+(\mathbf{A}|x) = \ &\text{trace}\{\boldsymbol{\Sigma}_0\} - \frac{\mathbf{A}'\boldsymbol{\Sigma}_0\mathbf{A}}{\mathbf{A}'\mathbf{A}} \\
&\times \left[1 - \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(\frac{\mathbf{A}'\mathbf{f}_{xx}(\omega)\mathbf{A}}{\mathbf{A}'\bar{\mathbf{f}}_{xx}\mathbf{A}}\right) d\omega\right\}\right], \quad (1)
\end{aligned}
$$

where $\bar{\mathbf{f}}_{xx} = (1/2\pi)\int_{-\pi}^{\pi} \mathbf{f}_{xx}(\omega)\, d\omega$. The denominator in the logarithmic term in Equation (3) cancels with $\mathbf{A}'\boldsymbol{\Sigma}_0\mathbf{A}$ to reduce to the earlier form given for $\text{FE}^+(\mathbf{A}|x)$. By Jensen's inequality, the term in square brackets lies in the interval [0, 1]. The lower extreme occurs when the spectrum is uniform; here $z(t)$ is totally unpredictable (at least by linear methods) and is similar to white noise. As the spectrum becomes more and more concentrated at a particular frequency, $z(t)$ is more predictable, and in the limit the term in square brackets converges to unity.

### 2.2 Formal Definition

Minimizing the forecast error upper bound $\text{FE}^+(\mathbf{A}|x)$ (1) with respect to $\mathbf{A}$ forces $\mathbf{A}'\boldsymbol{\Sigma}_0\mathbf{A}$ to be large (as in PCA) and the associated forcing function $z(t) = \mathbf{A}'x(t)$ to be highly predictable (as in POPS). Accordingly, the first PROP is defined as

$$
\mathbf{A}_1 = \underset{\mathbf{A}}{\text{argmin}}\{\text{FE}^+(\mathbf{A}|x)\}.
$$

Higher order PROPS are obtained in a recursive manner. Suppose that the first $j$ PROPS have been defined and let $\mathbf{A}^j$ be the matrix whose columns contain those vectors $\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_j\}$. Let $r(t)$ be the process associated with correcting $x(t)$ for the information already contained in the first $j$ PROPS; that is,

$$
r(t) = x(t) - E[x(t)|\mathbf{H}_t(\mathbf{A}^j)],
$$

where $\mathbf{H}_t(\mathbf{A}_j) = \{\mathbf{A}_j'x(s); s \leq t\}$. The $(j+1)$st PROP is defined as the first PROP (1) of the process $r(t)$,

$$
\mathbf{A}_{j+1} = \underset{\{\mathbf{A}:\mathbf{A}\perp\mathbf{A}^j\}}{\text{argmin}}\{\text{FE}^+(\mathbf{A}|r)\}. \quad (2)
$$

Note that the constraint that $\mathbf{A}_{j+1}$ be orthogonal to previous components is consistent with PCA. A total of $p$ PROPS can be defined in this manner. The $K$th order PROPS representation for $x(t)$ is

$$
x(t) = \sum_{j=1}^{K} \mathbf{A}_j z_j(t) + \eta^K(t),
$$

where $z_j(t) = \mathbf{A}_j'x(t)$ and $\eta^K(t) = x(t) - \sum_{j=1}^{K} \mathbf{A}_j z_j(t)$. The quantity $\sum_{j=1}^{K} \mathbf{A}_j z_j(t)$ is the least squares projection of $x(t)$ onto the column space of the first $K$ PROPS.

### 2.3 An Illustrative Example

Differences and similarities between the various decompositions can be highlighted by consideration of the situation in which $x(t)$ is assumed to follow a simple but pedagogically useful model,

$$
x(t) = \mathbf{U}z(t) + \boldsymbol{\varepsilon}(t).
$$

Here $z(t)$ is a scalar and $\boldsymbol{\varepsilon}(t)$ and $\mathbf{U}$ are vectors of length $p$. It is assumed that $z(t)$ and $\boldsymbol{\varepsilon}(t)$ are independent stationary processes. Assume that $\boldsymbol{\varepsilon}(t)$ is mean zero and temporally uncorrelated with the marginal covariance matrix $\boldsymbol{\Sigma}_\varepsilon$, and that $\mathbf{U}$ is a fixed unit vector. Here we have

$$
\begin{aligned}
\boldsymbol{\Sigma}_0 &= \mathbf{U}\mathbf{U}'\sigma_{zz}(0) + \boldsymbol{\Sigma}_\varepsilon, \\
\boldsymbol{\Sigma}_1 &= \mathbf{U}\mathbf{U}'\sigma_{zz}(1)
\end{aligned}
$$

and

$$
\mathbf{f}_{xx}(\omega) = \mathbf{U}\mathbf{U}'\mathbf{g}_{zz}(\omega) + \frac{1}{2\pi}\boldsymbol{\Sigma}_\varepsilon,
$$

where $\sigma_{zz}(k)$ is the lag-$k$ autocovariance of $z(t)$ and $\mathbf{g}_{zz}(\omega)$ is its spectral density.

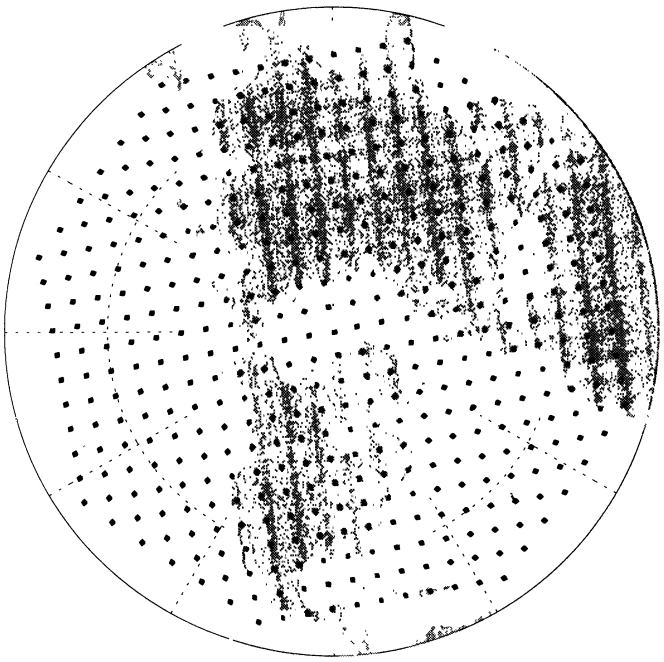

*Figure 1. The 445-Point Half-Resolution NMC Grid.*

Letting $\mathbf{H}^+(t) = \{(z(s), \boldsymbol{\varepsilon}(s)); s \le t\}$ be an ideal but unobservable history, the expected squared one-step-ahead forecast error for $x(t)$ is bounded *below* by

$$\begin{aligned} & E\|x(t+1) - E[x(t+1)|\mathbf{H}^+(t)]\|^2 \\ &= E\|\boldsymbol{\varepsilon}(t+1) + \mathbf{U}z(t+1) - E[\mathbf{U}z(t+1)|\mathbf{H}^+(t)]\|^2 \\ &= \text{trace}\{\boldsymbol{\Sigma}_\varepsilon\} + E\|\mathbf{U}z(t+1) - E[\mathbf{U}z(t+1)|\mathbf{H}^+(t)]\|^2 \\ &= \text{trace}\{\boldsymbol{\Sigma}_\varepsilon\} + 2\pi \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log \mathbf{g}_{zz}(\omega)\, d\omega\right\} \\ &\le \text{FE}^+(\mathbf{A}|x), \end{aligned}$$

provided that the logarithm of $\mathbf{g}_{zz}$ is integrable. From the definition for $\text{FE}^+(\mathbf{A}|x)$, we get that

$$\begin{aligned} & \text{FE}^+(\mathbf{A}|x) \\ &= \text{trace}\{\boldsymbol{\Sigma}_0\} - \mathbf{A}'\boldsymbol{\Sigma}_0\mathbf{A} \\ &\quad + 2\pi \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log(\mathbf{A}'\mathbf{f}_{xx}(\omega)\mathbf{A})\, d\omega\right\} \\ &= \sigma_{zz}(0) + \text{trace}\{\boldsymbol{\Sigma}_\varepsilon\} - \mathbf{A}'[\mathbf{U}\mathbf{U}'\sigma_{zz}(0) + \boldsymbol{\Sigma}_\varepsilon]\mathbf{A} \\ &\quad + 2\pi \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log(\mathbf{A}'\mathbf{f}_{xx}(\omega)\mathbf{A})\, d\omega\right\}. \end{aligned}$$

If $\boldsymbol{\Sigma}_\varepsilon$ is not of full rank and $\boldsymbol{\Sigma}_\varepsilon\mathbf{U} = 0$, then with $\mathbf{A} = \mathbf{U}$, we have

$$\text{FE}^+(\mathbf{A}|x) = \text{trace}\{\boldsymbol{\Sigma}_\varepsilon\} + 2\pi \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log \mathbf{g}_{zz}(\omega)\, d\omega\right\},$$

which, in view of the lower bound above, is the minimum value of the forecast error; that is, $\text{FE}^+ = \text{FE}$. Thus under these conditions, $\mathbf{A} = \mathbf{U}$ is seen to be the first PROP. It is easy to verify that $\mathbf{U}$ is an eigenvector of $\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1}$. In fact,

$$\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1}\mathbf{U} = \frac{\sigma_{zz}(1)}{\sigma_{zz}(0)}\,\mathbf{U}.$$

Thus if $|\sigma_{zz}(1)| > 0$, then the first POPS pattern of this model will be $\mathbf{U}$. However, it is not necessarily the case that the series $\mathbf{U}'x(t)$ follows an AR(1) model.

Situations in which PCA and POPS may fail to recover $\mathbf{U}$ are obtained as follows. Let $\lambda_1$ be the largest eigenvalue of $\boldsymbol{\Sigma}_\varepsilon$; then (a) the first *principal component* will fail to capture $\mathbf{U}$ whenever $\sigma_{zz}(0) < \lambda_1$; and (b) the first *principal oscillation pattern* may fail to coincide with $\mathbf{U}$ whenever $|\sigma_{zz}(1)| = 0$. All decompositions are the *same* whenever $\sigma_{zz}(0) > \lambda_1$ and $|\sigma_{zz}(1)| > 0$. On the other extreme, if $\sigma_{zz}(0) < \lambda_1$ and $|\sigma_{zz}(1)| = 0$, then both POPS and PCA can fail to extract $\mathbf{U}$ as the first component.

## 3. ESTIMATION METHODOLOGY

The approach here consists of replacing the covariance matrix $\boldsymbol{\Sigma}_0$ and the spectral density $\mathbf{f}_{xx}(\omega)$ in Equation (1) by sample estimates to obtain the first PROP. A similar approach is used for the higher order PROPS, with a least squares autoregression used to estimate the residual term in Equation (2). Suppose that the data consist of a finite record $\{x(t), t = 1, 2, \ldots T\}$. For simplicity, assume that $T$ is even and let $n = T/2$. The discrete Fourier transform of $x$ and corresponding periodogram at frequency $\omega_j = 2\pi j/T$ are

$$X_j^{(T)} = \frac{1}{\sqrt{T}} \sum_t x(t)e^{-i\omega_j t}$$

and

$$I_{xx}^{(T)}(\omega_j) = \frac{1}{2\pi}\, X_j^{(T)} X_j^{(T)*}$$

for $j = -(n+1), \ldots, n$. Smoothing the periodogram leads to the mean squared consistent estimator of the spectral density matrix of $x(t)$ (Brillinger 1981; Priestley 1987). Using a kernel approach, the spectral density estimate is

$$\mathbf{f}_{xx}^{(h)}(\omega) = \sum_j k_h(\omega - \omega_j) I_{xx}^{(T)}(\omega_j),$$

where $k_h(s) = (2\pi/Th)k(s/h), h > 0$, and $k(\cdot)$ is a kernel that integrates to unity and satisfies desired moment conditions (Priestley 1987; Silverman 1986). As $h \to 0, \mathbf{f}_{xx}^{(h)}$ reduces to the standard periodogram estimator of the spectral density: a sum of delta-functions at the frequencies $\omega_j$. Replacing $\boldsymbol{\Sigma}_0$ by the sample covariance $\hat{\boldsymbol{\Sigma}}_0$ and the spectral density matrix by $\mathbf{f}_{xx}^{(h)}$ and substituting into Equation (1) gives an estimation criterion: The first PROP $\hat{\mathbf{A}}_1$ is the maximizer of

$$\frac{\mathbf{A}'\hat{\boldsymbol{\Sigma}}_0\mathbf{A}}{\mathbf{A}'\mathbf{A}} \cdot \left[1 - \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(\frac{\mathbf{A}'\mathbf{f}_{xx}^{(h)}(\omega)\mathbf{A}}{\mathbf{A}'\bar{\mathbf{f}}_{xx}^{(h)}\mathbf{A}}\right) d\omega\right\}\right]. \tag{3}$$

In our numerical implementation the integral in the expression is approximated by a simple trapezoidal rule based

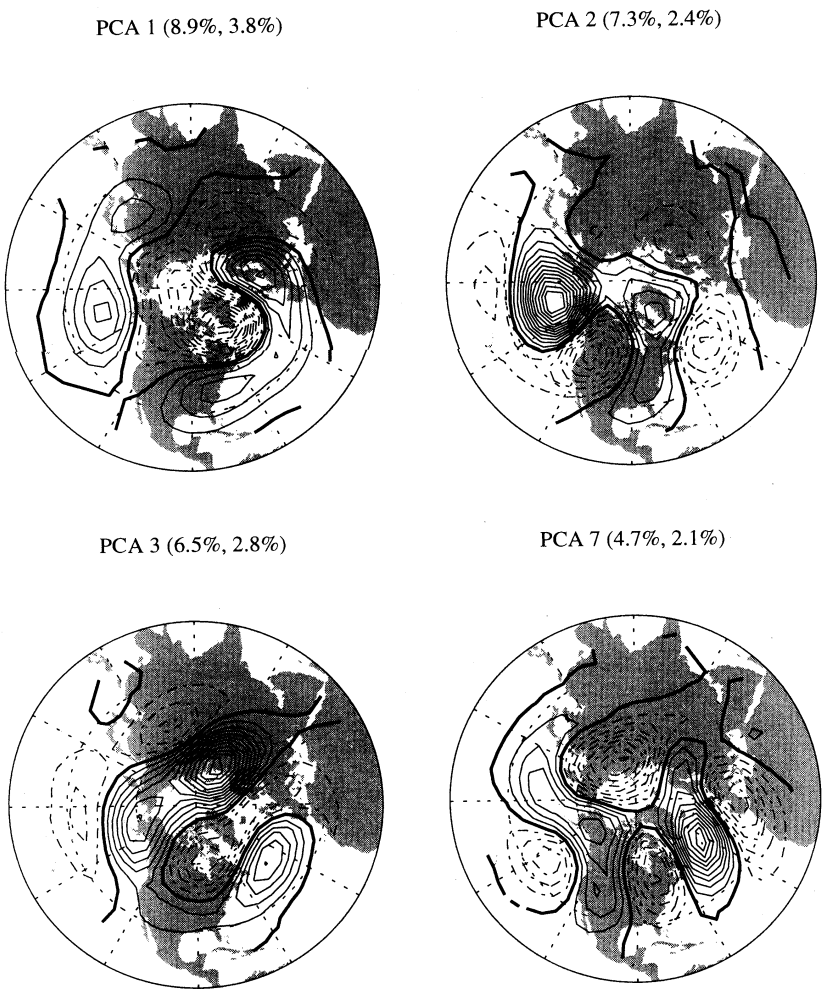


*Figure 2. The First Three and the Seventh Spatial Patterns Obtained by a Principal Component Analysis (PCA). The numbers above each figure are the percent of variance explained and the reduction in the forecast error (as in Sec. 3.1) obtained by including the pattern in the prediction model. The fields are dimensionless. The thick contour represents the zero-level, dashed contours are negative, and solid contours are positive.*

on the values of the integrand at the frequencies $\omega_j$ for $j = (-n+1), \ldots, n$. It should be emphasized that smoothing is a critical step in the estimation process. If $x(t)$ is real, then as $h \to 0$,

$$2\pi T \mathbf{A}' \mathbf{f}_{xx}^{(h)}(\omega_j)\mathbf{A} \propto \left[\mathbf{A}' \sum_j x(t)\cos(\omega_j t)\right]^2 + \left[\mathbf{A}' \sum_j x(t)\sin(\omega_j t)\right]^2 \equiv [\mathbf{A}'C_j]^2 + [\mathbf{A}'S_j]^2.$$

So without smoothing, $\mathbf{A}$ can be chosen to be orthogonal to $C_j$ and $S_j$, and then $\mathbf{A}'\mathbf{f}_{xx}^{(0)}(\omega_j)\mathbf{A} = 0$, leading to numerical instability for the integrand of the term in square brackets in Equation (3). Smoothing makes $\mathbf{f}_{xx}^{(h)}$ have full rank and thus overcomes this problem.

For higher order PROPS, an estimate of the process $r(t)$ is required. This is obtained from a lag-$L$ least squares autoregression,

$$\hat{r}(t) = x(t) - \sum_{l=0}^{L} \hat{\phi}_l z(t-l),$$

where, if we ignore all but the last $L$ terms, the coefficient matrices $\hat{\phi}_l$ can conveniently be chosen to minimize the sum of squares

$$\mathrm{RSS}[\phi] = \sum_{t=L+1}^{T} \left\| x(t) - \sum_{l=0}^{L} \phi_l z(t-l) \right\|^2$$

and $z(t) = (\hat{\mathbf{A}}^j)'x(t)$ with $\hat{\mathbf{A}}^j = \{\hat{\mathbf{A}}_1, \hat{\mathbf{A}}_2, \ldots \hat{\mathbf{A}}_j\}$. The orthogonality constraint requires $\hat{\mathbf{A}}^{j+1}$ to lie in the null space of $(\hat{\mathbf{A}}^j)'$. This is conveniently implemented via the $QR$ decomposition (Golub and Van Loan 1989). The $(j+1)$st estimated PROP is $\hat{\mathbf{A}}_{j+1} = \mathbf{Q}_2\hat{\alpha}$, where $\mathbf{Q}_2$ is the $p \times (p-j)$ matrix whose columns span the null space of $(\hat{\mathbf{A}}^j)'$ and $\hat{\alpha}$ is obtained as the unconstrained maximum of

$$\frac{\alpha'\mathbf{Q}_2'\hat{\mathbf{\Sigma}}_0\mathbf{Q}_2\alpha}{\alpha'\alpha} \cdot \left[1 - \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(\frac{\alpha' f_{\tilde{r}\tilde{r}}^{(h)}(\omega)\alpha}{\alpha' \bar{f}_{\tilde{r}\tilde{r}}^{(h)}\alpha}\right) d\omega\right\}\right], \tag{4}$$

where $f_{\tilde{r}\tilde{r}}^{(h)}$ is the spectral density of $\tilde{r}(t) = \mathbf{Q}_2'\hat{r}(t)$. Again, a trapezoidal rule is used to numerically evaluate the integral term.

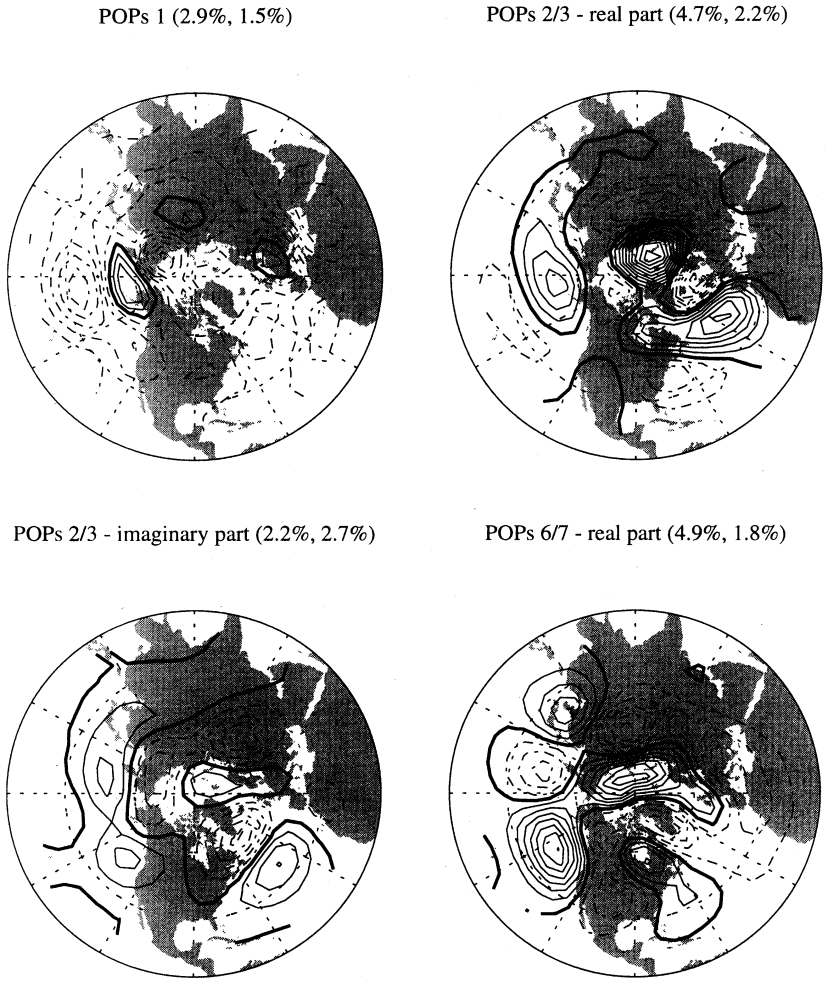


*Figure 3. The First Three and the Seventh Spatial Patterns Obtained by a Principal Oscillation Pattern Analysis (POPS). Percent variance and forecast errors are indicated as in Figure 2. Note that some of the components come in complex pairs.*

A steepest-descent algorithm is used to find the minimizers of the objective functions in Equations (3) and (4). The procedure is initialized in terms of the eigenvectors of $\hat{\mathbf{\Sigma}}_0$ or $\mathbf{Q}_2'\hat{\mathbf{\Sigma}}_0\mathbf{Q}_2$. For each eigenvector, the objective function is evaluated, and the eigenvector giving the minimum value is used as the initial guess. The steepest-descent algorithm requires only that the gradient of the objective function be computed. Explicit formulas for these gradients are used. In general, steepest-descent methods are more robust than more sophisticated Newton based methods, even though the local convergence characteristics of Newton methods are superior (Dennis and Schnabel 1983). The most time-consuming part of the computation is the calculation and storage of the spectral density matrix estimates. In view of this, the additional efficiency that one might achieve using Newton methods is likely to be negligible. The objective functions are likely not convex in $\alpha$—all maxima occur at least twice, because $\alpha$ and $-\alpha$ give identical values in Equation (4). Nevertheless, the steepest-descent procedure has not been found to be particularly sensitive to initialization. Indeed, in our experience with several hundred simulated data sets and the real data set in Section 4, the result of the algorithm found essentially the same value even when the procedure was initialized with a uniform starting guess for $\mathbf{A}_1$; that is, a vector all of whose components are equal to $1/\sqrt{p}$.

### 3.1 Predictive Evaluation of the PROPS Representation

A $K$th order PROPS representation for $x(t)$ is

$$x(t) = \sum_{j=1}^{K} \hat{\mathbf{A}}_j \hat{z}_j(t) + \eta^K(t),$$

where $\hat{z}_j(t) = \hat{\mathbf{A}}_j' x(t)$ and $\eta^K(t) = x(t) - \sum_{j=1}^{K} \hat{\mathbf{A}}_j \hat{z}_j(t)$. The predictive value of the PROPS representation can be evaluated in terms of an estimate of the one-step-ahead forecast error. To obtain an honest estimate, the data are split in two halves; the PROPS are estimated on the first half of the data, and a corresponding forecast rule is also estimated there. Then, this forecast rule is evaluated on the second half of the data, providing an estimate of the predictive performance of the model.

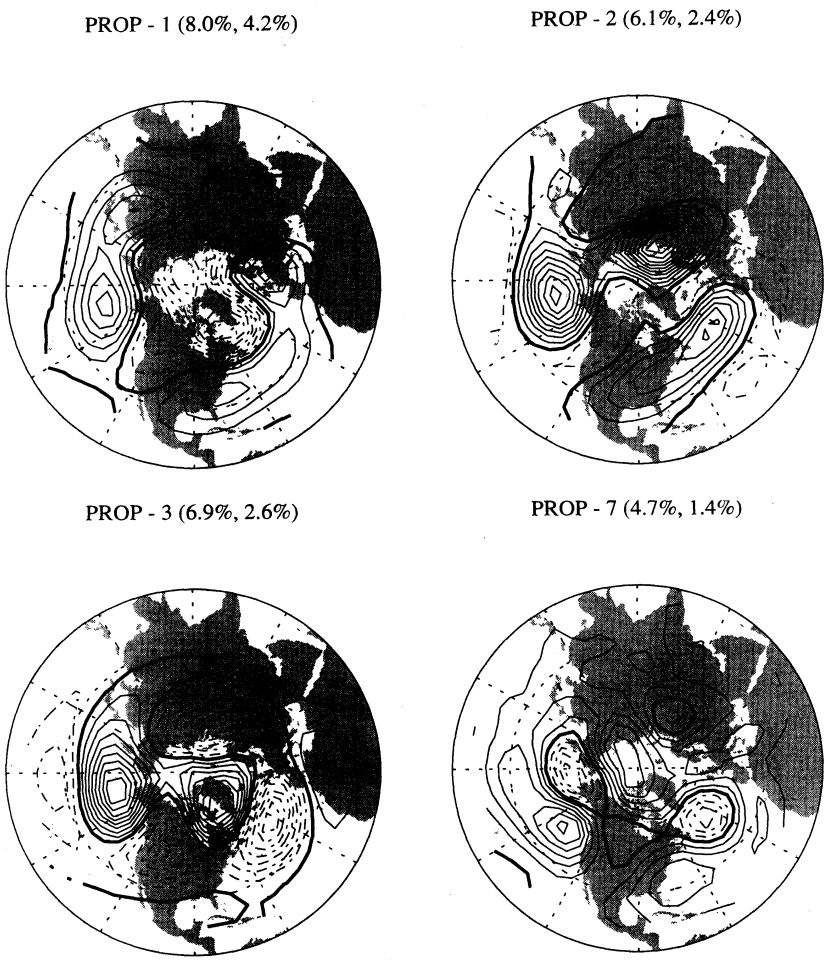


*Figure 4. The First Three and the Seventh Spatial Patterns Obtained by the Predictive Oscillation Pattern Analysis (PROPS). Percent variance and forecast errors are indicated as in Figure 2.*

In particular, we use multiple-response linear regression of the true field on lagged versions of the fitted PROPS representation of the field. The lags used were determined by backward elimination (see Kooperberg and O'Sullivan 1994 for details).

Two parameters that potentially influence the predictive value of a PROPS representation are the bandwidth ($h$) of the spectral density estimate and the order of the lag ($L$) used in defining the process $\hat{r}(t)$. In our computations we fixed these to $h = 20$ and $L = 2$. However, calculations not reported here suggest that the influence of these parameters is modest. For a further discussion of the bandwidth, see the Appendix.

## 4. APPLICATION TO THE GEOPOTENTIAL HEIGHT ANOMALY FIELD

The geopotential (500 hPa) height is the single atmospheric quantity that best summarizes the low-frequency dynamics of the atmosphere (Wallace and Gutzler 1981). As such, the geopotential height field has been studied extensively in the atmospheric sciences, both by itself and in relation to other variables, such as the sea surface temperature (see, e.g., Dickson and Namias 1976, Horel 1981, Kushnir and Wallace 1989, and Wallace et al. 1992). We used a 47-year climatological record (January 1946–May 1989) of Northern Hemisphere extra-tropical geopotential (500 hPa) height, as projected onto a 445-point half-resolution National Meteorology Center (NMC) grid based on daily operational analyses from the NMC. The NMC grid is a point octagonal superimposed on a polar stereographic projection of the Northern Hemisphere, extending to 20 degrees north. Figure 1 (p. 1488) shows the location of the grid points. Note that the projection in all of our figures shows equal areas on the globe as equal areas in the figure. The NMC grid is one of the commonly used grids for studying large-scale dynamics of the atmosphere. An advantage of the NMC grids is that each point corresponds to the same area on the ground. (If different points represented different areas, then one could argue that a weighting procedure was needed.) The actual gridding is not very important, as the patterns we are looking for are so large scale in space that they should be similar on any grid.

Previously, methods such as PCA and POPS have been applied to these data to derive insight into climatological weather patterns (Wallace et al. 1992). Similar to these analyses, the data were then time-averaged to produce 5-day averages. Not surprisingly, there are strong seasonal trends in the data. The series was made approximately stationary by removing the mean field and the first three harmonics of

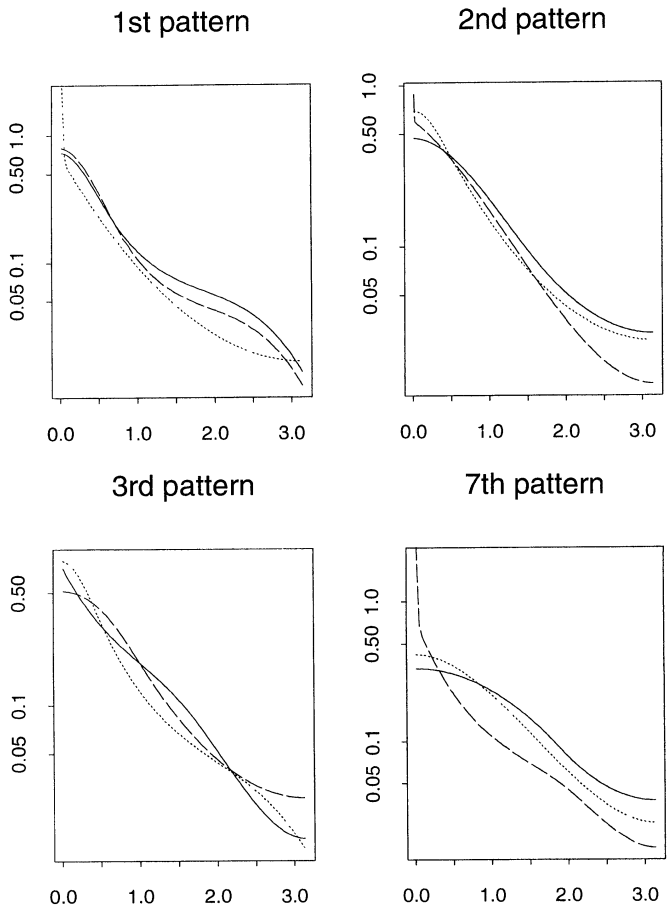


*Figure 5. Spectra of the Temporal Forcing Functions (Normalized to Have Unit Variance) Associated With the Spatial Scales of Variations Shown in Figures 2, 3, and 4. The solid line represents PCA; the dotted line, POPS; the dashed line, PROPS.*

the annual cycle, separately for every grid point. This operation was carried out prior to time-averaging to produce the 5-day averages. Five-day averages are of interest, because 5 days seems to be the time lag for which statistical methods may be able to contribute to improved forecasting. In particular, 1- and 2-day forecasts based on physical models have little room left for improvement, whereas forecasts beyond, say, 8 days are only marginally correlated with the actual field 8 days later.

In this article we examine the complete series of 3,467 5-day averages. The corrected data set is referred to as the anomaly field. The data were divided into two parts, with the first 1,700 records used for estimation and the remaining half used for forecast evaluation (following Kooperberg and O'Sullivan (1994). In Figure 2 we show the first three patterns and the seventh pattern identified by PCA on the first half of the data. The percentages above the figures are the percentage of variance explained by the pattern within the first half of the data and the forecast error obtained on the second half of the data (see Sec. 3.1). It is assuring that these PCA patterns appear to change little when the length of the (5-day) averaging procedure is changed or when the second half of the data is used. In particular, patterns for the second half of the data are extremely similar to those shown here. When 3-day averages were used, the same patterns as in Figure 2 were still clearly recognizable, although the first two patterns seemed to be swapped.

*Table 1. One-Step Prediction Variances for the Spectral Densities in Figure 5*

| Method | PCA | POPS | PROPS |
|---|---|---|---|
| Pattern 1 | .56 | .34 | .48 |
| Pattern 2 | .63 | .47 | .47 |
| Pattern 3 | .55 | .56 | .59 |
| Pattern 7 | .75 | .69 | .47 |

The first and second PCA's are recognized as combinations of the North American Oscillation (NAO) pattern and the Pacific North American (PNA) pattern, which are well documented in the climatology literature (Dickson and Namias 1976). The patterns indicate a connection between the behavior of the atmosphere over the North Pacific and North America. The NAO pattern manifests itself as a sharp peak and trough in the North Atlantic (as in the first PCA), whereas the PNA pattern is recognizable by the sequence of alternating high and low values over the North Pacific and North America. These patterns appear in many different analyses, using various different griddings of the atmosphere.

To make computations manageable, the 445-dimensional data were projected onto the first 50 principal components (four of which are shown in Fig. 2). These components explain in excess of 95% of the marginal variance. After computing the PROPS on the $p = 50$ dimensional time series of principal component scores, the spatial patterns and PROPS representation of the original 445-dimensional field are defined as linear combinations of the principal component vectors of the original. Other computations, not reported here, suggest that the value of $p$ has a negligible effect on the patterns estimated by PROPS.

For the POPS analysis, the same projected data was used. However, for POPS such a projection is needed as a method to regularize the patterns (Von Storch et al. 1988). Similar projections have been commonly used in other analyses of such data sets (Barnett and Preisendorfer 1987; Wallace et al. 1992).

Figures 3 and 4 show a subset of spatial patterns identified by the POPS and PROPS techniques. As in Figure 2, the percent variance explained and the percent forecast error obtained on the second half of the data are also indicated. The first PROP is very similar to the first PCA. The second PROP is different from the second PCA (it looks more like the third PCA), but the PNA pattern is still readily recognizable. The first POPS pattern is not immediately interpretable in terms of known climatological patterns; neither is it as spatially coherent (smooth) as the PNA or NAO pattern. The other three patterns obtained by all methods appear spatially coherent but are quite different from each other. The closest two patterns are the second PROP and the third PCA, both of which show a high degree of correlation in the region associated with the Gulf Stream.

Figure 5 shows spectral density estimates of the temporal forcing functions associated with each pattern in Figures 2–4 (on a logarithmic scale). These estimates were computed using an automated spectral density estimator developed

by Kooperberg, Stone, and Truong (1995). In these graphs the forcing functions have been normalized to have unit variance so that all spectra integrate to the same value. This allows a comparison of one-step prediction variances,

$$\sigma_\varepsilon^2 = 2\pi \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log h(\omega)\, d\omega \right\} \tag{5}$$

(see Priestley 1987, eq. 10.1.41). We evaluated the integral in (5) numerically for each of the 12 spectral densities in Figure 5; the results are summarized in Table 1.

It can be seen from Table 1 that the first POPS pattern is the most predictable pattern (as is also evident from the fact that the spectrum corresponding to the first POPS is the most peaked pattern). However, because the POPS pattern, which explains 2.9% of the variance in the field, is much less correlated with the field than the first PCA pattern, which explains 8.9% of the variance in the field, and the first PROPS pattern, which explains 8.0%, the reduction in forecast error achieved by the first POPS pattern (1.5%) is still considerably smaller than the reduction in forecast error achieved by the first PROPS (4.0%) and the PCA (3.8%) patterns. The first PROPS pattern is sufficiently more predictable than the first PCA pattern, so that, irrespective of the fact that the first PCA is more correlated with the field, the first PROPS achieves a higher reduction in forecast error.

The PROPS patterns have a richer structure than that of a simple AR(1) model. Fitting AR models to the forcing functions lead to models with orders between 6 and 23, using an Akaike information criterion (AIC) model selection criterion (Priestley 1987). The deviation from the AR(1) structure potentially undermines the ability to appeal to the elegant interpretation of POPS representations developed by Von Storch et al. (1988). The forecast error performance of the different methods are compared in Figure 6. The graph displays

$$\text{Error}[j] = \frac{\sum_t \|x(t) - \hat{x}_j(t)\|^2}{\sum_t \|x(t)\|^2},$$

where $\hat{x}_j(t)$ is the *optimal* linear forecast of $x(t)$ based on the information contained in the history of the first $j$ components of the representation under study (PCA, POPS, or PROPS); that is, $H_{t-1}(\mathbf{A}^j) = \{(\mathbf{A}^j)'x(s); s < t\}$. Note that $\hat{x}(t)$ is computed on the first half of the data and the error is computed on the second half of the data (see Sec. 3.1). In terms of this performance criterion, the PROPS representation is seen to be more favorable. It is important to realize that the forecast error for all three methods is not particularly good in absolute terms. The main reasons for this are that although some of the earlier patterns (e.g., the first and second PROPS patterns and the first POPS pattern) by themselves are fairly predictable, these patterns explain only a small percentage of the variance in the field.

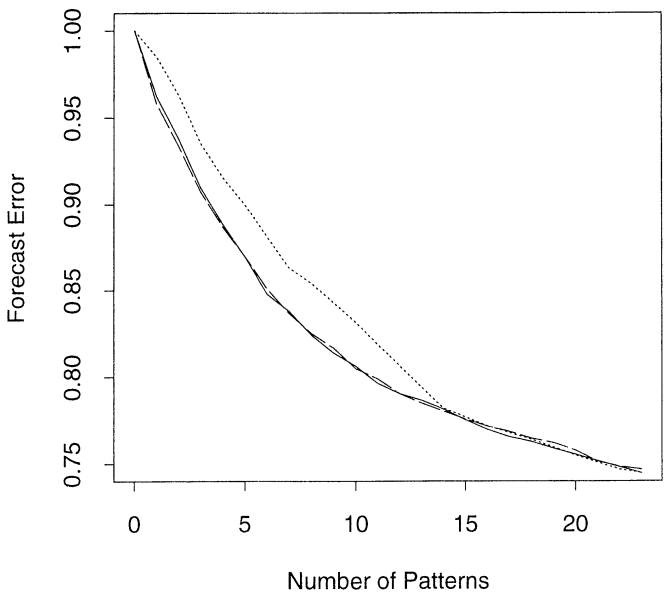


*Figure 6. Forecast Error Performance of the Alternative Decomposition Methods as a Function of the Number of Patterns Included. There is only a slight improvement in the forecast error after 25 components; hence these are not included. The solid line represents PCA; the dotted line, POPS; the dashed line, PROPS.*

PCA, POPS, and PROPS analyses yield distinctive representations for the geopotential height field. Although the forecast error criterion would favor the PROPS representation, the choice is not very clear cut. In any event, the PROPS analysis provides alternative complementary representation, which may be useful in developing an understanding of the climatology associated with this important meteorological variable.

## 5. DISCUSSION

We have introduced a new approach to constructing spatial-temporal decompositions of random fields. The method explicitly retains desirable characteristics of both PCA and POPS. We have numerically implemented the method and applied it to a climatological data set, comparing the results to alternative decompositions.

Various directions for future development of this methodology might be considered. The incorporation of smoothing constraints on the spatial scales of variation might be approached by the addition of an appropriate roughness penalty to the objective function. Incorporation of phase effects into the PROPS representations, along the lines suggested by Bürger (1993) for the POPS analysis, is also of interest. Another generalization would be to develop analogs of canonical correlations enabling the representation of coupling between pairs of climatological fields (Von Storch et al. 1988). More sophisticated approaches to constructing spatial-temporal representations for a field might also be considered. A natural approach would be to consider representations in terms of a state-space or latent variable model (Priestley 1987); for example, $x(t) = \mathbf{A}z(t) + \varepsilon(t)$. With stationarity assumptions for $z$ and $\varepsilon$, the spectral density matrix of the field is

$$f_{xx}(\omega) = \mathbf{A}\mathbf{g}_{zz}(\omega)\mathbf{A}' = h_{\varepsilon\varepsilon},$$

where $\mathbf{g}_{zz}$ is the spectral density matrix of the $K$-dimensional latent process $z(t)$, $\mathbf{A}$ is a $p \times K$ loading matrix, and $h_{\varepsilon\varepsilon} = (1/2\pi)\mathbf{\Sigma}_\varepsilon$ is the spectral density of the error term $\varepsilon(t)$. Thus in the frequency domain one might con-

sider estimation of this factor analysis–type model using an appropriate Whittle likelihood (Whittle 1962). Implementation of this approach poses a number of computational as well as some theoretical statistical problems. In particular, because $\mathbf{A}$ and $\mathbf{g}_{zz}$ are more closely related, it is no longer clear whether $\mathbf{A}$ can be estimated at a fully efficient semiparametric rate.

Rapidly developing global databases from atmospheric science and oceanography with high-resolution spatial and temporal sampling, like the geopotential height field considered in this article, could well provide a valuable frame of reference and source of stimulation for the further development of applied statistical methodology related to the analysis of random fields. The recent report of the Panel on Statistics and Oceanography (Chelton 1994; Panel on Statistics and Oceanography 1994; Von Storch 1994) indicates many problems on which atmospheric scientists, oceanographers, and statisticians could cooperate.

## APPENDIX: SOME ASYMPTOTIC CONSIDERATIONS

From the study of semiparametric estimation problems, one knows that there are situations in which the rate of estimation of a finite-dimensional parameter is adversely affected by the estimation of a nonparametric nuisance parameter (Bickel, Klaassen, Ritov, and Wellner 1993; Engle, Granger, and Rice 1986; Speckman 1988). The PROPS estimation procedure involves using a kernel smoother to approximate the spectral density. As a result, it is not clear if the PROPS estimates converge to their true values at a parametric rate. To examine this, we consider the estimation of the first PROP and by one-step linearization consider convergence characteristics when the length of the series $T$ increases but the dimension of the field $p$ remains fixed. Up to the linearization, the analysis shows that the proposed estimator converges at a regular parametric rate. Furthermore, the asymptotic mean squared error does not appear to be sensitive to the choice of the form of the kernel. It would be of interest to study whether these properties also hold under different situations; for example, when $p$ and $T$ increase simultaneously. We do not pursue this issue here.

Let $\mathbf{H}[-\pi,\pi]$ be a class of functions that map the interval $[-\pi,\pi]$ to complex Hermitian matrices of dimension $p$. Consider the real-valued functional $l : R^p \times \mathbf{H}[-\pi,\pi] \to R$,

$$l(a,f) = \frac{\int_{-\pi}^{\pi} a' f(\omega) a \, d\omega}{a'a} \times \left[1 - \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(\frac{a' f(\omega) a}{a' \bar{f} a}\right) d\omega\right\}\right], \quad \text{(A.1)}$$

where $\bar{f} = (1/2\pi)\int_{-\pi}^{\pi} f(\omega)\, d\omega$. This functional is of interest because the first PROP is its maximum with $f = \mathbf{f}_{xx}^{(h)}$, whereas the true value is defined as the maximizer with $f = \mathbf{f}_{xx}$. Let $a[f]$ denote a maximizer of the functional for a general $f$. Although $a[f]$ is not unique ($l(a,f) = l(-a,f)$), in the neighborhood of $a[f]$ when $f$ is chosen sufficiently regular, there is a unique root of the score equation $\partial_a l(a[f], f) = 0$. A first-order Taylor series expansion for the vector-valued function $b(s) = a[\mathbf{f}_{xx} + s(\mathbf{f}_{xx}^{(h)} - \mathbf{f}_{xx})]$ gives $b(1) \approx b(0) + \dot{b}(0)$, where $\dot{b}(0) = (db/ds)(s)|_{s=0}$. Thus

$$a[\mathbf{f}_{xx}^{(h)}] \approx a[\mathbf{f}_{xx}] + \dot{a}[\mathbf{f}_{xx} + s(\mathbf{f}_{xx}^{(h)} - \mathbf{f}_{xx})]_{s=0} \equiv a[\mathbf{f}_{xx}] + \partial_f a[\mathbf{f}_{xx}](\mathbf{f}_{xx}^{(h)} - \mathbf{f}_{xx}), \quad \text{(A.2)}$$

where the second term is the directional derivative of $a[\cdot]$ at $\mathbf{f}_{xx}$ in the direction of the function $(\mathbf{f}_{xx}^{(h)} - \mathbf{f}_{xx})$. Differentiation of the score equation $\partial_a l(a[f + s\delta], f + s\delta) = 0$ with respect to $s$ gives the formula

$$\partial_f a[\mathbf{f}_{xx}](\mathbf{f}_{xx}^{(h)} - \mathbf{f}_{xx}) = -\partial_a^2 l(a[\mathbf{f}_{xx}], \mathbf{f}_{xx})^{-1} \partial_a(\partial_f l(a[\mathbf{f}_{xx}], \mathbf{f}_{xx})(\mathbf{f}_{xx}^{(h)} - \mathbf{f}_{xx})).$$

This is a consequence of the implicit function theorem (Rall 1969). Here we have used the fact that the order of the differentiation with respect to $a$ and $f$ can be interchanged, which follows from the regularity of the functional $l$. Thus, to within the first-order approximation of Equation (A.1), the expected squared error of the first PROP in a direction defined by a vector $\mathbf{u}$ is

$$E[\mathbf{u}'(a[\mathbf{f}_{xx}^{(h)}] - a[\mathbf{f}_{xx}])]^2 = E[\mathbf{v}'\partial_a(\partial_f l(a[\mathbf{f}_{xx}], \mathbf{f}_{xx})(\mathbf{f}_{xx}^{(h)} - \mathbf{f}_{xx}))]^2,$$

where $\mathbf{v} = \partial_a^2 l(a[\mathbf{f}_{xx}], \mathbf{f}_{xx})^{-1}\mathbf{u}$. From Equation (5), an expression for the directional derivative of $l$ in the direction of a function $\delta$ is

$$\begin{aligned}\partial_f l(a,f)\delta \;=\; & \frac{\int_{-\pi}^{\pi} a'\,\delta(\omega) a \, d\omega}{a'a} \\ & \times \left[1 - \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(\frac{a' f(\omega) a}{a' \bar{f} a}\right) d\omega\right\}\right] \\ & - \frac{\int_{-\pi}^{\pi} a' f(\omega) a \, d\omega}{a'a} \\ & \times \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(\frac{a' f(\omega) a}{a' \bar{f} a}\right) d\omega\right\} \\ & \times \left[\frac{1}{2\pi}\left(\int_{-\pi}^{\pi} \frac{a'\delta(\omega)a}{a' f(\omega) a} d\omega - \frac{a'\bar{\delta} a}{a'\bar{f} a}\right)\right]. \end{aligned} \quad \text{(A.3)}$$

The directional derivative of $\partial_f l(a,f)\delta$ with respect to $a$ in the direction of the vector $\mathbf{v}$ is $\mathbf{v}'\partial_a\partial_f l(a,f)\delta$. In view of Equation (A.2), it is easy to appreciate that the expression for this quantity is rather elaborate. It suffices to consider some representative terms,

$$g_0(a,f) \cdot \int_{-\pi}^{\pi} [a'\,\delta(\omega) a] \, d\omega$$

and

$$g_1(a,f) \cdot \int_{-\pi}^{\pi} [\mathbf{v}'\,\delta(\omega) a + a'\,\delta(\omega)\mathbf{v}] g_2(\omega | a, f(\omega)) \, d\omega,$$

where $g_0(a,f)$ and $g_1(a,f)$ are independent of $\delta$ and $\mathbf{v}$ and $g_2(\omega|a,f(\omega))$ is either the function $1/(a'f(\omega)a)$ or the constant function. Focusing on these terms, it can be seen that the asymptotic squared error characteristic of the estimator is determined by the behavior of random variables of the form

$$\beta^{(h)} = \int_{-\pi}^{\pi} \boldsymbol{\gamma}'[\mathbf{f}_{xx}^{(h)}(\omega) - \mathbf{f}_{xx}(\omega)]\boldsymbol{\eta}\theta(\omega)\, d\omega,$$

where $\boldsymbol{\eta}$ and $\boldsymbol{\gamma}$ are fixed vectors and $\theta(\omega) = g_2(\omega | a[\mathbf{f}_{xx}], \mathbf{f}_{xx}(\omega))$ is a fixed real-valued function possibly depending on $1/(a'\mathbf{f}_{xx}(\omega)a)$.

Separating $\beta^{(h)}$ into its systematic and stochastic components gives

$$\beta^{(h)} = \int_{-\pi}^{\pi} \boldsymbol{\gamma}'[E\mathbf{f}_{xx}^{(h)}(\omega) - \mathbf{f}_{xx}(\omega)]\boldsymbol{\eta}\theta(\omega)\,d\omega + \int_{-\pi}^{\pi} \boldsymbol{\gamma}'[\mathbf{f}_{xx}^{(h)}(\omega) - E\mathbf{f}_{xx}^{(h)}(\omega)]\boldsymbol{\eta}\theta(\omega)\,d\omega$$
$$\equiv \beta_s^{(h)} + \beta_r^{(h)}.$$

If the kernel function integrates to unity and has zero first moment and second moment equal to unity (Silverman 1986), and if $\theta(\omega)$ has a continuous second derivative, then as $h \to 0$,

$$\begin{aligned}
&\int_{-\pi}^{\pi} \frac{1}{h}\, k\left(\frac{\omega - \omega_j}{h}\right)\theta(\omega)\,d\omega \\
&\quad \approx \int_{-\pi}^{\pi} \left[\theta(\omega_j) + \dot{\theta}(\omega_j)(\omega - \omega_j) + \ddot{\theta}(\omega_j)\,\frac{(\omega - \omega_j)^2}{2}\right] \\
&\qquad \times \frac{1}{h}\, k\left(\frac{\omega - \omega_j}{h}\right) d\omega \\
&\quad = \theta(\omega_j) + \ddot{\theta}(\omega_j)\,\frac{h^2}{2}.
\end{aligned}$$

Substituting into the expression for the systematic component of $\beta^{(h)}$ and using the approximation $EI_{xx}(\omega_j) \approx \mathbf{f}_{xx}(\omega_j)$ (Brillinger 1981; Priestley 1987) gives

$$\begin{aligned}
\beta_s^{(h)} &\approx \frac{2\pi}{T} \sum_{j=-n+1}^{n} \boldsymbol{\gamma}'\mathbf{f}_{xx}(\omega_j)\boldsymbol{\eta}\theta(\omega_j) \\
&\quad - \int_{-\pi}^{\pi} \boldsymbol{\gamma}'\mathbf{f}_{xx}(\omega_j)\boldsymbol{\eta}\theta(\omega)\,d\omega \\
&\quad + \frac{2\pi}{T} \sum_{j=-n+1}^{n} \boldsymbol{\gamma}'\mathbf{f}_{xx}(\omega_j)\boldsymbol{\eta}\ddot{\theta}(\omega_j)\,\frac{h^2}{2} \\
&= O(T^{-1}) + O(h^2)
\end{aligned}$$

as $T \to \infty$ and $h \to 0$. The order of the approximations can be verified provided that the spectral density is smooth; for example, has a continuous second derivative (Brillinger 1981; Priestley 1987).

For the stochastic component, we have

$$\beta_r^{(h)} \approx \frac{2\pi}{T} \sum_{j=-n+1}^{n} \left[\theta(\omega_j) + \ddot{\theta}(\omega_j)\,\frac{h^2}{2}\right] \boldsymbol{\gamma}'[I_{xx}(\omega_j) - \mathbf{f}_{xx}(\omega_j)]\boldsymbol{\eta}.$$

The second-order behavior of the periodogram has been well studied (Brillinger 1981; Priestley 1987), and so one has under the usual regularity conditions that

$$\begin{aligned}
E[\beta_r^{(h)}]^2 &= \left(\frac{2\pi}{T}\right)^2 \sum_{j=-n+1}^{n} [\theta(\omega_j)^2\mathbf{v}(\omega_j) + O(h^4)] \\
&= \frac{2\pi}{T}\left[\int_{-\pi}^{\pi} \theta(\omega)^2\mathbf{v}(\omega)\,d\omega + O(T^{-1}) + O(h^4)\right],
\end{aligned}$$

where $\mathbf{v}(\omega_j) = E\|\boldsymbol{\gamma}'[W_j - EW_j]\boldsymbol{\eta}\|^2$ and $W_j$ follows a complex Wishart distribution with 1 degree of freedom and scale matrix $\mathbf{f}_{xx}(\omega_j)$. Therefore, if $h = cT^s$ for $s < -\frac{1}{4}$, then $E[\beta^{(h)}]^2 = \mathbf{A}T^{-1}$, where $\mathbf{A}$ is independent of the kernel.

Note that under second-order smoothness assumptions for the spectral density, the familiar $O(T^{-4/5})$ rate of convergence for the mean integrated squared error of the smoothed periodogram is achieved with $h = O(T^{-1/5})$ (Brillinger 1981; Priestley 1987). Thus it is apparent that the parametric rate of convergence for the first PROP is obtained by using an *undersmoothed* estimate of the spectral density. Note, however, that the spectral density estimate can still be mean squared consistent; for example, a bandwidth on the order of $h = O(T^{-1/3})$ is fine.

## A.1 NUMERICAL ILLUSTRATION

A rigorous development of the foregoing heuristics would be of some interest, but we shall not pursue that here. Instead, we end this section by reporting on a simple numerical study reinforcing the main conclusion from the first-order linearization argument. We consider a model in which $x(t)$ has $p = 7$ components independent of each other. Components 3–7 are white Gaussian processes with mean zero and variance 1.0. The second component is also Gaussian white-noise but with variance 4.0. The first component is a fourth-order moving average process defined by

$$x_1(t) = -.3e(t-1) - .6e(t-2) - .3e(t-3) + .6e(t-4) + e(t),$$

where $e(t)$ is a standard Gaussian white-noise process. From Section 2.3, the first PROP (and POP) of this model is the unit vector $\mathbf{A} = (1, 0, 0, 0, 0, 0, 0)'$. Repeated data sets were generated from the model, with lengths $T$ ranging from 256 to 4,096. A total of 40 replications for each value of $T$ were considered. The first PROP was computed for each data set and compared to the true value, after appropriate adjustment of the sign.

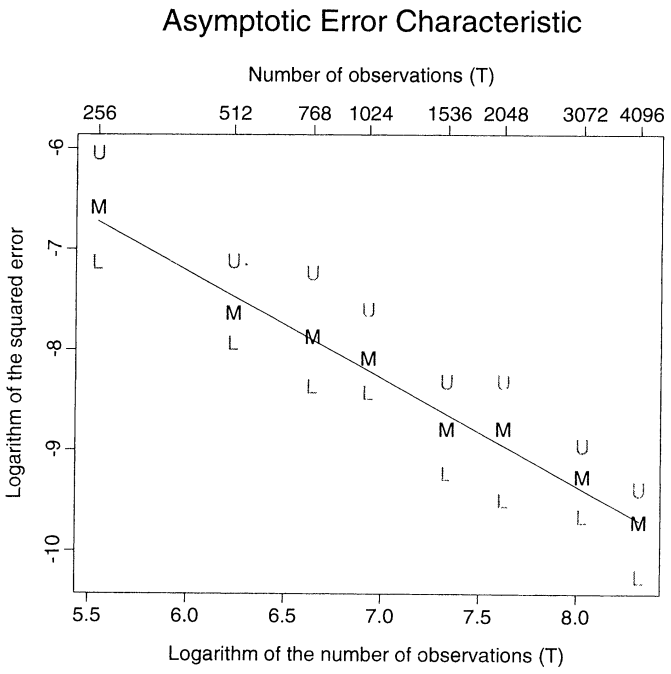


*Figure A.1. The Relation Between the Squared Error of the First PROP and the Number of Observations (see Sec. 5.1). Forty replications were obtained at each value of T; the median (M), upper (U), and lower quantiles (L) are shown. The line is a least squares fit to the logarithm of the squared error as a linear function of the logarithm of T. A parametric rate of convergence would have a slope of −1.0.*

Figure A.1 shows the squared error difference between the true and estimated PROP as a function of $T$. From the theoretical development we expect a linear logarithmic relation with slope of $-1.0$. Figure A.1 confirms this intuition. The estimated relation has a slope of $-1.06$ with a standard error of .04, consistent with the theoretical analysis. A fixed bandwidth was used in this analysis. Interestingly, varying the bandwidth did not appear to affect either the slope or the intercept.

*[Received August 1994. Revised February 1996.]*

## REFERENCES


Barnett, T. P., and Preisendorfer, R. W. (1987), "Origins and Levels of Monthly and Seasonal Forecast Skill for United States Surface Air Temperature Determined by Canonical Correlations Analysis," *Monthly Weather Review*, 115, 1825–1850.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993), *Efficient and Adaptive Estimation for Semiparametric Models*, Baltimore: Johns Hopkins University Press.

Bretherton, C. S., Smith, C., and Wallace, J. M. (1992), "An Intercomparison of Methods for Finding Coupled Patterns in Climate Data," *Journal of Climate*, 5, 541–560.

Brillinger, D. R. (1981), *Time Series: Data Analysis and Theory*, San Francisco: Holden-Day.

Bürger, G. (1993), "Complex Principal Oscillation Pattern Analysis," *Journal of Climate*, 6, 1972–1986.

Chelton, D. B. (1994), "Physical Oceanography: A Brief Overview for Statisticians," *Statistical Science*, 9, 150–166.

Dennis, J. E., and Schanbel, R. B. (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Englewood Cliffs, NJ: Prentice-Hall.

Dickson, R. R., and Namias, J. (1976), "North American Influences on the Circulation and Climate of the North Atlantic Sector," *Monthly Weather Review*, 104, 1255–1265.

Engle, R. F., Granger, C. W. J., and Rice, J. (1986), "Semiparametric Estimates of the Relation Between Weather and Electricity Sales," *Journal of the American Statistical Association*, 81, 310–320.

Golub, G. H., and Van Loan, C. F. (1989), *Matrix Computations*, Baltimore: Johns Hopkins University Press.

Hasselmann, K. (1988), "PIPs and POPs: The Reduction of Complex Dynamical Systems Using Principal Interaction and Oscillation Patterns," *Journal of Geophysical Research*, 93 D9, 11,015–11,021.

Holton, J. R. (1992), *An Introduction to Dynamic Meteorology* (3rd ed.), San Diego: Academic Press.

Horel, J. D. (1981), "A Rotated Principle Component Analysis of the Interannual Variability of the Northern Hemisphere 500 mb Height Field," *Monthly Weather Review*, 109, 2080–2092.

Kooperberg, C., and O'Sullivan, F. (1994), "The Use of a Statistical Forecast Criterion to Evaluate Alternative Empirical Spatial Oscillation Pattern Decomposition Methods in Climatological Fields," Technical Report 276, University of Washington, Dept. of Statistics.

Kooperberg, C., Stone, C. J., and Truong, Y. K. (1995), "Logspline Estimation of a Possibly Mixed Spectral Distribution," *Journal of Time Series Analysis*, 16, 359–388.

Kushnir, Y., and Wallace, J. M. (1989), "Low-Frequency Variability in the Northern Hemisphere Winter: Geographical Distribution, Structure and Time-Scale Dependence," *Journal of the Atmospheric Sciences*, 46, 3122–3142.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979), *Multivariate Analysis*, London: Academic Press.

Panel on Statistics and Oceanography (1994), "Report on Statistics and Physical Oceanography" (with discussion), *Statistical Science*, 9, 167–221.

Priestley, M. B. (1987), *Spectral Analysis and Time Series*, London: Academic Press.

Rall, L. B. (1969), *Computational Solution of Nonlinear Operator Equations*, New York: Wiley.

Rao, C. R. (1973), *Linear Statistical Inference*, New York: Wiley.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Speckman, P. (1988), "Kernel Smoothing in Partial Linear Models," *Journal of the Royal Statistical Society*, Ser. B, 50, 413–436.

Von Storch, H. (1994), "Contribution to the Discussion of the Report on Statistics and Physical Oceanography," *Statistical Science*, 9, 215–221.

Von Storch, H., Bruns, T., Fischer-Bruns, I., and Hasselmann, K. (1988), "Principal Oscillation Pattern Analysis of the 30- to 60-Day Oscillation in General Circulation Model Equatorial Troposphere," *Journal of Geophysical Research*, 93 D9, 11,022–11,036.

Wallace, J. M., and Gutzler, D. S. (1981), "Teleconnections in the Geopotential Height Field During the Northern Hemisphere Winter," *Monthly Weather Review*, 109, 785–812.

Wallace, J. M., and Hobbs, P. V. (1977), *Atmospheric Sciences: An Introductory Survey*, New York: Academic Press.

Wallace, J. M., Smith, C., and Bretherton, C. S. (1992), "Singular Value Decomposition of Wintertime Sea Surface Temperature and 500 mb Height Anomalies," *Journal of Climate*, 5, 561–576.

Whittle, P. (1962), "Gaussian Estimation in Stationary Time Series," *Bulletin of the International Institute of Statistics*, 39, 105–129.