An Intercomparison of Methods for Finding Coupled Patterns in Climate Data

CHRISTOPHER S. BRETHERTON, CATHERINE SMITH, AND JOHN M. WALLACE

Department of Atmospheric Sciences, University of Washington, Seattle, Washington

(Manuscript received 23 July 1990, in final form 28 May 1991)

ABSTRACT

This paper introduces a conceptual framework for comparing methods that isolate important coupled modes of variability between time series of two fields. Four specific methods are compared: principal component analysis with the fields combined (CPCA), canonical correlation analysis (CCA) and a variant of CCA proposed by Barnett and Preisendorfer (BP), principal component analysis of one single field followed by correlation of its component amplitudes with the second field (SFPCA), and singular value decomposition of the covariance matrix between the two fields (SVD). SVD and CPCA are easier to implement than BP, and do not involve user-chosen parameters. All methods are applied to a simple but geophysically relevant model system and their ability to detect a coupled signal is compared as parameters such as the number of points in each field, the number of samples in the time domain, and the signal-to-noise ratio are varied.

In datasets involving geophysical fields, the number of sampling times may not be much larger than the number of observing locations or grid points for each field. In a model system with these characteristics, CPCA usually extracted the coupled pattern somewhat more accurately than SVD, BP, and SFPCA, since the patterns it yielded exhibit smaller sampling variability; SVD and BP gave quite similar results; and CCA was uncompetitive due to a high sampling variability unless the coupled signal was highly localized. The coupled modes derived from CPCA and SFPCA exhibit an undesirable mean bias toward the leading EOFs of the individual fields; in fact, for small signal-to-noise ratios these methods may identify a coupled signal that is similar to a dominant EOF of one of the fields but is completely orthogonal to the true coupled signal within that field. For longer time series, or in situations where the coupled signal does not resemble the EOFs of the individual fields, these biases can make SVD and BP substantially superior to CPCA.

1. Introduction

The analysis of the relationships within and among datasets involving large grid point (or station) arrays and time series can be done in many different ways. Simple methods of analysis like compositing and correlation based on selected "reference grid points" or indices are easy to perform, but they involve subjective decisions about the choice of reference time series. Matrix operations offer the possibility of a more objective definition of the structures in such datasets.

One such technique applicable to measurements of a single field at many locations is called principal component analysis (PCA). First proposed by Pearson (1902), PCA identifies linear transformations of the dataset that concentrate as much of the variance as possible into a small number of variables. A related technique, factor analysis, was introduced around the same time as Pearson's work by Spearman (1904a,b) and extended by Hotelling (1935, 1936), who formulated mathematical equations for defining unique solutions. These techniques were first used extensively in the social sciences. The first applications to meteorology were by Fukuoka (1951), Lorenz (1956), Holmstrom (1963), and Obukhov (1960). The lucid exposition of PCA by Kutzbach (1967) was instrumental in promoting the use of this technique in climate research.

Our focus will be on providing a unified conceptual framework for techniques that isolate important *coupled* modes of variability between time series of two fields. Several approaches have been applied to geophysical data; we will concentrate on four of these.

Kutzbach (1967) pointed out that two or more field variables can be combined in the same PCA to document the relationships between the fields. We will refer to this technique as combined PCA (CPCA). A second method is to correlate the expansion coefficients of the EOFs of one of the fields with the second field to obtain correlation maps for this EOF with the second field; this method will be called single-field–based PCA, or SFPCA, (e.g., Wallace et al. 1990). A third widely used technique is canonical correlation analysis (CCA), which is designed to identify the linear combinations of variables in one field that are most strongly correlated with linear combinations of variables in another field. This method was first devised by Hotelling (1935, 1936) and has been widely used in the social sciences

Corresponding author address: John M. Wallace, University of Washington, Department of Atmospheric Sciences, AK-40, Seattle, WA 98195.

since the 1960s. A typical application is discussed in the textbook of Cooley and Lohnes (1971). A group of students answers a set of questions designed to provide information on their occupational interests and their personality traits. There are many more student respondents than there are items on the questionnaire. A principal component analysis, based upon the complete dataset consisting of both types of questions [analogous to the methodology used by Kutzbach (1967)], yields factors related to maturity, interest in manual labor, intellectual ability, etc. Because the questions concerning interests and personality traits are not treated separately, this method does not necessarily yield any information concerning the relationship between the answers to these two groups of questions. CCA of the same dataset yields related pairs of factors from the two groups of questions. The first factor, for example, shows that an interest in middle-class occupations tends to be shared by those with middle-class values; the second relates an outgoing personality with an interest in sports and outdoor occupations, etc.

The first meteorological application of CCA was by Glahn (1968), who used it in the context of statistical weather prediction. It has been subsequently used by Barnett (1981, 1983) and Barnett and Preisendorfer (1987) to relate patterns of seasonal mean SST anomalies over the Pacific to surface air temperature anomalies over the United States the following season; by Nicholls (1987) to study teleconnections related to the Southern Oscillation; by Déqué and Servain (1989) to study lead/lag relationships between monthly mean SST and 700-mb height patterns over the Atlantic sector; by Metz (1989) to study the relationship between low-frequency fluctuations in the 500-mb height field and the associated patterns of forcing by the high-frequency fluctuations; and by Graham (1990) to identify related simultaneous and lag-correlation patterns in tropical Pacific SST and the Northern Hemisphere 700-mb height field. Preisendorfer (1988) devoted a chapter of his monograph on principal component analysis in meteorology to a discussion of CCA. Barnett and Preisendorfer (1987) suggested that filtering the data for each field by projecting it onto a subset of the EOFs of that field can make CCA less susceptible to sampling fluctuations due to short time series. We call this variant of CCA the BP method.

A fourth method, singular value decomposition, or SVD, of the covariance matrix between two fields, has not been widely used in meteorology but is the simplest of the four methods to perform, is easily interpreted, and (we will show in this paper) yields results similar to the more elaborate BP method. SVD is a fundamental matrix operation, a generalization of the diagonalization procedure that is performed in PCA to matrices that are not square or symmetric. It is available in almost every linear algebra or statistics software package, takes approximately the same amount of computer resources as diagonalizing a square matrix, and has a wide variety of applications in fields such as inverse theory. Like CCA, SVD isolates linear combinations of variables within two fields that tend to be linearly related to one another. The optimality criterion in SVD is somewhat different than that used in CCA.

SVD was first used in a meteorological context by Prohaska (1976) to document the simultaneous relationships between monthly mean surface air temperature over the United States and hemispheric sea level pressure patterns. It has been used by Lanzante (1984) to study the relationship between seasonal mean extratropical SST and 700-mb height anomalies and by Dynmnikov and Filin (1985) to document the relationship between extratropical SST fluctuations and atmospheric heating fields during the year of the GARP Global Weather Experiment.

This paper introduces a conceptual framework in which these methods are compared (section 2) and summarizes their salient features (section 3). In section 4, a simple but geophysically relevant model system is introduced. We compare the quantitative performance of the methods in isolating a "coupled" signal as we vary parameters such as the spatial localization of the coupled signal, the number of sampling times, the number of grid points in each field, and the ratio of the coupled signal amplitude to uncoupled variability. In a companion paper (Wallace et al. 1992), we have compared the performance of SVD, CCA, and CPCA in documenting the relationships between SST anomalies and anomalies in the atmospheric 500-mb geopotential height field over the North Pacific.

2. Nomenclature

In this section, some concepts and nomenclature that are common to all methods of identifying correlated patterns from observations of two time-dependent fields. Consider a "left" data field $\mathbf{s}(\mathbf{x}, t)$. In the companion paper, this field consists of SST anomaly values at N_s grid points x_i for T observation times. Consider another "right" data field $\mathbf{z}(\mathbf{y}, t)$, consisting of 500-mb height anomaly values $z_j(t)$ at N_z (possibly different) grid points for the same T observation times. We will always denote time series by the notation (t) and vectors (whose components do not depend on time or are all observed at the same time) with boldface type. In the companion paper, each time series is *normalized* to unit standard deviation, but for this analysis we will require only that the time series both have zero mean. Note that any analysis of the coupling between the fields will give meaningful results only if there are significant correlations between some of the $s_i(t)$ and some of the $z_j(t)$.

The data time series $\mathbf{s}(t)$ and $\mathbf{z}(t)$ at each of the grid points can each be expanded in terms of a set of N vectors, called *patterns*, which will depend upon the analysis method used:

$$\mathbf{s}(t) \leftarrow \tilde{\mathbf{s}}(t) \equiv \sum_{k=1}^{N} a_k(t)\mathbf{p}_k, \quad (1a)$$

$$\mathbf{z}(t) \leftarrow \tilde{\mathbf{z}}(t) \equiv \sum_{k=1}^{N} b_k(t)\mathbf{q}_k. \quad (1b)$$

The time series $a_k(t)$ and $b_k(t)$ are called *expansion coefficients* and are also method dependent. With a limited number $N < T$ of patterns $\mathbf{p}_k$ and $\mathbf{q}_k$, we may not be able to recover the exact gridpoint time series no matter how the expansion coefficients are chosen, hence the notation "$\leftarrow$" instead of "=." In general, it is required that the "synthetic" time series $\tilde{\mathbf{s}}(t)$ and $\tilde{\mathbf{z}}(t)$ be "as close as possible" (again, the meaning of this varies with the method) to the actual time series. For all methods discussed, the expansion coefficients are calculated as weighted linear combinations of the gridpoint data:

$$a_k(t) = \sum_{i=1}^{N_s} u_{ik}s_i(t) = \mathbf{u}_k^T\mathbf{s}(t), \quad (2a)$$

$$b_k(t) = \sum_{j=1}^{N_z} v_{jk}z_j(t) = \mathbf{v}_k^T\mathbf{z}(t). \quad (2b)$$

The vectors $\mathbf{u}_k$ and $\mathbf{v}_k$ will be called *weight vectors.*

Together, each pair of patterns, the corresponding pair of weight vectors, and the pair of expansion coefficients define a *mode.* Throughout the text, $\mathbf{p}_k$ and $\mathbf{q}_k$ will always be used for the patterns, and $\mathbf{u}_k$ and $\mathbf{v}_k$ for the weight vectors associated with a mode, regardless of the analysis method being discussed. The convention is that individual grid points of the left and right fields will always be subscripted with i and j, respectively, and that individual modes of the two fields will always be subscripted with k (and l if more than one mode is referenced).

All of the methods can be discussed in terms of the covariances between gridpoint observations of the fields. Let $\langle f(t) \rangle$ denote the time average of a time series $f(t)$ over the T observation times. It is convenient to define the single-field covariance matrices

$$C_{ss} = \langle \mathbf{s}(t)\mathbf{s}^T(t) \rangle \quad (N_s \times N_s), \quad (3a)$$

$$C_{zz} = \langle \mathbf{z}(t)\mathbf{z}^T(t) \rangle \quad (N_z \times N_z), \quad (3b)$$

the cross-covariance matrix between the fields

$$C_{sz} = \langle \mathbf{s}(t)\mathbf{z}^T(t) \rangle \quad (N_s \times N_z), \quad (3c)$$

and the combined covariance matrix

$$C_{s+z} = \langle (\mathbf{s}(t)|\mathbf{z}(t))(\mathbf{s}(t)|\mathbf{z}(t))^T \rangle \quad ((N_s + N_z) \times (N_s + N_z)). \quad (3d)$$

For several of the methods PCA must be performed on one or both of the fields or their combination. The following notation is used:

eigenvalues and eigenvectors of C_{ss}: $(\lambda_m, \mathbf{e}_m)$,
normalized left PCA expansion coefficients: $\alpha_m(t) = \lambda_m^{-1/2}\mathbf{s}^T(t)\mathbf{e}_m$,
eigenvalues and eigenvectors of C_{zz}: $(\kappa_n, \mathbf{f}_n)$,
normalized right PCA expansion coefficients: $\beta_n(t) = \kappa_n^{-1/2}\mathbf{z}^T(t)\mathbf{f}_n$.

For PCA, the eigenvectors are spatially orthonormal and the normalized expansion coefficients are uncorrelated and have unit variance. It will sometimes prove useful to express the time series for one or both fields in terms of its *PC (principal component) basis,* in which the EOFs (eigenvectors of the variance matrix) of this field are used to define the spatial structures and their associated normalized time series of expansion coefficients specify the temporal evolution.

a. Correlation and covariance maps

Using the expansion coefficients $a_k(t)$ and $b_k(t)$ from any method, two types of correlation maps can be generated. Let $r[f(t),g(t)]$ denote the correlation coefficient between two time series $f(t)$ and $g(t)$ (if one of the two time series is a vector, we mean the vector of correlation coefficients between its component time series and the other time series). The kth left *homogeneous correlation map* is defined to be the vector $r[\mathbf{s}(t),a_k(t)]$ of correlations between the gridpoint values of the left field and the kth left expansion coefficient. Similarly, the kth left *heterogeneous correlation map* is the vector of correlation coefficients between the gridpoint values of the left field and the kth expansion coefficient of the *right* field. The left homogeneous correlation map is a useful indicator of the geographic localization of the covarying part of the s field, while the left heterogeneous correlation map indicates how well the grid points in the left field can be predicted from the kth right expansion coefficient (derived from the right field). All of this discussion applies equally well if "left" and "right" are interchanged.

In a parallel manner, one may also define homogeneous and heterogeneous *covariance maps* of the covariance of an expansion coefficient with gridpoint values of a field. If the gridpoint data are normalized, the covariance map is proportional to the correlation map. If in addition the expansion coefficient has variance one, the two maps are identical.

b. Comparative measure of explained covariance

It is useful to have a measure that indicates how successful each method has been in explaining the observed covariance matrix C_{sz} between the fields using a small number N of the dominant modes. In this section, we introduce one such measure, called the *cumulative squared covariance fraction,* or CSCF. Let $\tilde{C}_N$ be a synthetic covariance matrix constructed as explained below using just these N modes. The cumu-

lative squared covariance fraction explained by these modes is defined as

$$\mathrm{CSCF}_N = 1 - \frac{\|C_{sz} - \tilde{C}_N\|_F^2}{\|C_{sz}\|_F^2}. \quad (4)$$

Here the Frobenius matrix norm of C (denoted by subscript F) is defined to be the square root of the total amount of squared covariance summed over all entries in C:

$$\|C\|_F^2 \equiv \sum_i^{N_s} \sum_j^{N_z} C_{ij}^2. \quad (5)$$

The CSCF will be close to unity if the synthetic covariance matrix is a good fit to the actual covariance matrix and less otherwise. The squared covariance fraction SCF_k explained by a *single* mode k can be defined similarly by using just mode k to construct $\tilde{C}$.

To calculate $\tilde{C}_N$, a synthetic time series $\tilde{\mathbf{s}}_N(t)$ of left field gridpoint values of the left field using just the first N modes. The synthetic time series is the sum over the modes of the product of the left expansion coefficient with the left spatial pattern for that mode:

$$\tilde{\mathbf{s}}_N(t) = \sum_{k=1}^{N} a_k(t)\mathbf{p}_k. \quad (6a)$$

A similar synthetic time series $\tilde{\mathbf{z}}_N(t)$ is found for the right field:

$$\tilde{\mathbf{z}}_N(t) = \sum_{k=1}^{N} b_k(t)\mathbf{q}_k. \quad (6b)$$

The synthetic covariance matrix is then defined as

$$\tilde{C}_N = \langle \tilde{\mathbf{s}}_N(t)\tilde{\mathbf{z}}_N^T(t) \rangle. \quad (7)$$

3. Salient features of some analysis methods for coupled fields

Table 1 summarizes the analysis methods that are compared in this paper. Each method satisfies some optimality criterion (although for PCA, the optimality criterion is based on one field only, rather than any measure of correlation between the fields). In this section we will elaborate on Table 1 and will discuss for each method how left and right patterns are deduced and the relevance of the optimality criterion to pattern finding. The formulas in Table 1 for the synthetic covariance matrix for each method can be derived using the orthogonality relations for the given method in (6).

a. Singular value decomposition

Singular value decomposition (SVD) is a fundamental matrix operation that can be thought of as an extension to rectangular matrices of the diagonalization of a square symmetric matrix. As shown below, singular value decomposition of the cross-covariance matrix identifies, from two data fields, pairs of spatial patterns that explain as much as possible of the mean-squared temporal covariance between the two fields. The abbreviation SVD will be used to denote both the general matrix operation and the specific application to C_{sz}; the meaning should be clear from the context.

In SVD a spatially orthonormal set of patterns, which will be specified presently, and a synthetic time series $\tilde{\mathbf{s}}(t)$, formed as a linear combination of patterns will be considered to be as close as possible to the true time series $\mathbf{s}(t)$ of a field if at all observation times $\mathbf{s}(t) - \tilde{\mathbf{s}}(t)$ is orthogonal to all N patterns. Then the expansion coefficients are just generalized Fourier coefficients and the weight vectors are just the patterns themselves:

$$a_k(t) = \mathbf{p}_k^T\mathbf{s}(t), \quad b_k(t) = \mathbf{q}_k^T\mathbf{z}(t),$$
$$\mathbf{u}_k = \mathbf{p}_k, \quad \mathbf{v}_k = \mathbf{q}_k. \quad (8)$$

Because the patterns are spatially orthonormal, the left expansion coefficients can be thought of as the projections of the vector $\mathbf{s}(t)$ on the left patterns, and similarly for the right field.

The "leading" patterns $\mathbf{p}_1$ and $\mathbf{q}_1$ are chosen as follows: The projection $a_1(t)$ of $\mathbf{s}$ on $\mathbf{p}_1$ has the maximum covariance with the projection $b_1(t)$ of $\mathbf{z}(t)$ on $\mathbf{q}_1$. Successive pairs $(\mathbf{p}_k, \mathbf{q}_k)$ are chosen in exactly the same way with the added condition that $\mathbf{p}_k$ is orthogonal to $\mathbf{p}_1, \ldots, \mathbf{p}_{k-1}$, and $\mathbf{q}_k$ is orthogonal to $\mathbf{q}_1, \ldots, \mathbf{q}_{k-1}$.

The covariance of $a_1(t)$ and $b_1(t)$ can be written using (1c, d) and (2) as:

$$\langle a_1(t), b_1(t) \rangle = \mathbf{p}_1^T C_{sz}\mathbf{q}_1 = \max. \quad (9)$$

The choice of $\mathbf{p}_1$ and $\mathbf{q}_1$ that will maximize this covariance is deduced from the singular value decomposition of C_{sz}, whose properties are discussed by Strang (1988, pp. 443–452). Here, some important properties of the SVD are summarized for application to our problem.

(S1) Any $N_s \times N_z$ matrix C can be decomposed uniquely as follows:

$$C = \sum_{k=1}^{R} \sigma_k \mathbf{l}_k \mathbf{r}_k^T, \quad R \leqslant \min(N_s, N_z), \quad (10)$$

where the $\mathbf{l}_k$ are an orthonormal set of R vectors of length N_z called the *left singular vectors,* the $\mathbf{r}_k$ are an orthonormal set of R vectors of length N_z called the *right singular vectors,* and the σ_k are nonnegative numbers called the *singular values,* ordered such that $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r$ and R is the rank of C.

(S2) $C^T\mathbf{l}_k = \sigma_k\mathbf{r}_k$, $C\mathbf{r}_k = \sigma_k\mathbf{l}_k$.

(S3) The σ_k^2 are the R nonzero eigenvalues of C^TC and CC^T. The remaining eigenvalues of these two matrices are zero. For a covariance matrix taken from T observations, there are at most $T - 1$ nonzero singular values (the rank of C is no larger than $T - 1$). The rank is $T - 1$ rather than T because one degree of freedom is lost when the sample mean is subtracted from each observation to calculate covariances.

TABLE 1. A comparison of methods for determining connected patterns in two fields.

	Method									
	SVD		"Classical" CCA		BP		CPCA		LPCA	
Optimality property	Maximizes covariance between left and right spatially normalized weighted sums. Synthetic covariance matrix is the best possible fit to C_{sz} for a given number of modes.		Maximizes correlation between left and right weighted sums.		Maximizes correlation between left and right weighted sums after left data are projected onto leading N_1 left EOFs and right data are projected onto leading N_2 right EOFs.(§3)		Maximizes variance explained by a weighted sum of elements in both fields.		Maximizes variance in the left field explained by the left weight vector.	
Operations/results	SVD (C_{sz}) R singular values σ_k(§1) Left singular vectors $\mathbf{l}_k[N_s]$ Right singular vectors $\mathbf{r}_k[N_z]$		SVD ($C_{ss}^{-1/2}C_{sz}C_{zz}^{-1/2}$)(§2) R singular values $\tilde{\rho}_k$(§1) Left singular vectors $\tilde{\mathbf{l}}_k[N_s]$ Right singular vectors $\tilde{\mathbf{r}}_k[N_z]$		SVD ($C\alpha\beta$)$[N_1 \times N_2]$(§4) $\hat{R}$ singular values $\hat{\rho}_k$ Left singular vectors $\hat{\mathbf{l}}_k[N_1]$ Right singular vectors $\hat{\mathbf{r}}_k[N_2]$		PCA (C_{s+z}) $N_s + N_z$ eigenvalues η_k, Eigenvectors $\hat{\mathbf{g}}_k = (\hat{\mathbf{e}}_k \mid \hat{\mathbf{f}}_k)N_s + N_z$ Normalized expansion coefficients $\hat{\alpha}_k(t) = \eta_k^{-1/2}\hat{\mathbf{g}}_k \cdot (\mathbf{S}(t) \mid \mathbf{Z}(t))$		PCA (C_{ss}) N_s eigenvalues λ_k, Eigenvectors $\mathbf{e}_k[N_s]$, Normalized expansion coefficients $\alpha_k(t) = \lambda_k^{-1/2}\mathbf{e}_k \cdot \mathbf{S}(t)$	
	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right
Weight vectors $\mathbf{u}_k$, $\mathbf{v}_k$	$\mathbf{l}_k$	$\mathbf{r}_k$	$C_{ss}^{1/2}\tilde{\mathbf{l}}_k$	$C_{zz}^{1/2}\tilde{\mathbf{r}}_k$	$E\Lambda^{-1/2}\hat{\mathbf{l}}_k$(§5)	$FM^{-1/2}\hat{\mathbf{l}}_k$	$\eta_k^{-1/2}\hat{\mathbf{g}}_k$(§6)		$\lambda_k^{-1/2}\mathbf{e}_k$	(§6)
Patterns $\mathbf{p}_k$, $\mathbf{q}_k$	$\mathbf{l}_k$	$\mathbf{r}_k$	$C_{ss}\mathbf{u}_k$	$C_{zz}\mathbf{v}_k$	$E\Lambda^{1/2}\hat{\mathbf{l}}_k$	$FM^{1/2}\hat{\mathbf{l}}_k$	$\eta_k^{1/2}\hat{\mathbf{e}}_k$	$\eta_k^{1/2}\hat{\mathbf{f}}_k$	$\lambda_k^{1/2}\mathbf{e}_k$	$\lambda_k^{1/2}C_{sz}^T\mathbf{e}_k$
Expansion coefficients $a_k(t)$, $b_k(t)$	$l_k \cdot \mathbf{S}(t)$	$\mathbf{r}_k \cdot \mathbf{Z}(t)$	$\mathbf{u}_k \cdot \mathbf{S}(t)$	$\mathbf{v}_k \cdot \mathbf{Z}(t)$	$\hat{\mathbf{l}}_k \cdot \alpha(t)$	$\hat{\mathbf{r}}_k \cdot \beta(t)$	$\hat{\alpha}_k(t)$	$\hat{\alpha}_k(t)$	$\alpha_k(t)$	$\alpha_k(t)$
Homogeneous covariance maps $\langle \mathbf{S}, a_k \rangle$, $\langle \mathbf{Z}, b_k \rangle$	$C_{ss}\mathbf{l}_k$	$C_{zz}\mathbf{r}_k$	$\mathbf{p}_k$	$\mathbf{q}_k$	$\mathbf{p}_k$	$\mathbf{q}_k$	$\mathbf{p}_k$	$\mathbf{q}_k$	$\mathbf{p}_k$	$\mathbf{q}_k$
Heterogeneous covariance maps $\langle \mathbf{S}, b_k \rangle$, $\langle \mathbf{Z}, a_k \rangle$	$\sigma_k\mathbf{p}_k$	$\sigma_k\mathbf{q}_k$	$\tilde{\rho}_k\mathbf{p}_k$	$\tilde{\rho}_k\mathbf{q}_k$	$\hat{\rho}_k\mathbf{p}_k$	$\hat{\rho}_k\mathbf{q}_k$	$\mathbf{p}_k$	$\mathbf{q}_k$	$\mathbf{p}_k$	$\mathbf{q}_k$
Orthogonality relations	$\langle a_k, b_l \rangle = \sigma_k\delta_{kl}$ $\langle a_k, b_l \rangle, \langle b_k, b_l \rangle \neq \delta_{kl}$ $\mathbf{l}_k \cdot \mathbf{l}_l = \delta_{kl}$ $\mathbf{r}_k \cdot \mathbf{r}_l = \delta_{kl}$ $C_{sz}^T\mathbf{l}_k = \sigma_k\mathbf{r}_k$ $C_{sz}\mathbf{r}_k = \sigma_k\mathbf{l}_k$		$\langle a_k, a_l \rangle = \langle b_k, b_l \rangle = \delta_{kl}$ $\langle a_k, b_l \rangle = \tilde{\rho}_k\delta_{kl}$ $\mathbf{p}_k \cdot \mathbf{p}_l$, $\mathbf{q}_k \cdot \mathbf{q}_l \neq \delta_{kl}$		Same as classical CCA with $\tilde{\rho}_k$ replaced by $\hat{\rho}_k$.		$\langle \hat{\alpha}_k, \hat{\alpha}_l \rangle = \delta_{kl}$ $\hat{\mathbf{e}}_k \cdot \hat{\mathbf{e}}_l + \hat{\mathbf{f}}_k \cdot \hat{\mathbf{f}}_l = \hat{\mathbf{g}}_k \cdot \mathbf{g}_l = \delta_{kl}$		$\langle \alpha_k, \alpha_l \rangle = \delta_{kl}$ $\mathbf{e}_k \cdot \mathbf{e}_l = \delta_{kl}$	
Synthetic covariance matrix for N modes	$\Sigma_{k=1}^N \sigma_k\mathbf{p}_k\mathbf{q}_k^T$		$\Sigma_{k=1}^N \tilde{\rho}_k\mathbf{p}_k\mathbf{q}_k^T$		$\Sigma_{k=1}^N \hat{\rho}_k\mathbf{p}_k\mathbf{q}_k^T$		$\Sigma_{k=1}^N \mathbf{p}_k\mathbf{q}_k^T$		$\Sigma_{k=1}^N \mathbf{p}_k\mathbf{q}_k^T$	

(§0) Square brackets denote dimensions.
(§1) $R \leq \min(N_s, N_z)$ is rank of C_{sz}.
(§2) C_{ss} and C_{zz} must be invertible. This requires that $T > \max(N_s, N_z) - 1$.
(§3) N_1 and N_2 are user chosen. One choice is to keep the minimum number of EOFs of each field that explain a threshold (e.g., 70%) fraction of the variance of that field.
(§4) Here $\alpha(t) = (\alpha_1, \cdots, \alpha_{N_1})$, an N_1-vector of the leading N_1 expansion coefficients in PCA (C_{ss}). Also, $\hat{R} \leq \min(N_1, N_2)$ is the rank of $C\alpha\beta \equiv \langle \alpha\beta^T \rangle$.
(§5) $E = (\mathbf{e}_1 \mid \cdots \mid \mathbf{e}_{N_1})$, $[N_s \times N_1]$; $\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_{N_1} \end{pmatrix}[N_1 \times N_1]$; $F = (\mathbf{f}_1 \mid \cdots \mid \mathbf{f}_N)$, $[N_s \times N]$; $M = \begin{pmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_{N_2} \end{pmatrix}[N_2 \times N_2]$.
(§6) For CPCA there is one weight vector weighting both fields. For LPCA, the only expansion coefficient $a_k(t)$ does not depend on the right field.

(S3) The $\mathbf{l}_k$ are the eigenvectors of CC^T with nonzero eigenvalues, and the $\mathbf{r}_k$ are the eigenvectors of C^TC with nonzero eigenvalues.

(S5) For a square symmetric matrix, the left and right singular vectors are both equal to the eigenvectors and the singular values are the absolute values of the eigenvalues.

(S6) The square of the Frobenius matrix norm of C is the sum of the squares of its singular values:

$$\|C\|_F^2 = \sum_{k=1}^{R} \sigma_k^2. \tag{11}$$

The SVD can be computed with an efficient, numerically stable algorithm available in most linear algebra software packages (such as LINPACK and MatLab). For a full rank matrix C with $R = N_z \leqslant N_s$, the method takes $O(N_sN_z^2)$ floating point calculations (flops) (Golub and Van Loan 1983, p. 175). A straightforward but numerically less attractive method of obtaining the singular values used by Prohaska (1976) and Lanzante (1984) is to use (S3) and compute the singular values and right singular vectors as the eigenvalues and eigenvectors of the matrix CC^T; one can then find the left singular vectors from (S2). The advantages of the former method, as discussed by Golub and Reinsch (1970), are that it is more numerically stable and requires only a fraction of the number of flops.

With these properties in mind, the spatial patterns $\mathbf{p}_1$ and $\mathbf{q}_1$ we now expand in the bases of left and right singular vectors, respectively:

$$\mathbf{p}_1 = \sum_{m=1}^{N_s} \mu_m \mathbf{l}_m, \tag{12a}$$

$$\mathbf{q}_1 = \sum_{n=1}^{N_z} \eta_n \mathbf{r}_n. \tag{12b}$$

Taking the inner product of (12a) with itself, using the orthonormality of the $\mathbf{l}$, and noting that $|\mathbf{p}_1| = 1$, we deduced that $|\mu| = 1$. Similarly $|\eta| = 1$. Substituting (6) and (7) into (5), we find that

$$\begin{aligned}\langle a_1(t), b_1(t)\rangle &= \sum_{m=1}^{N_s}\sum_{n=1}^{N_z}\mu_m\eta_n\sum_{p=1}^{r}\sigma_p\mathbf{l}_m^T\mathbf{l}_p\mathbf{r}_p^T\mathbf{r}_n \\ &= \sum_{p=1}^{r}\mu_p\eta_p\sigma_p \leqslant \sigma_1\Big|\sum_{p=1}^{r}\mu_p\eta_p\Big| \\ &\leqslant \sigma_1|\mu\|\eta| = \sigma_1. \end{aligned}\tag{13}$$

This relationship becomes an equality only when $\mu_1 = \eta_1 = 1$ and all other coefficients μ_p, η_p are zero; that is, when $\mathbf{p}_1 = \mathbf{l}_1$, $\mathbf{q}_1 = \mathbf{r}_1$. Hence, the maximum covariance is equal to the largest singular value σ_1 and is obtained by projecting the left field $\mathbf{s}(t)$ onto the first left singular vector and the right field $\mathbf{z}(t)$ onto the first right singular vector.

Similarly, the subsequent pairs of patterns that explain the maximum amount of covariance of the two fields, subject to the constraint that they be orthogonal to the previous patterns, are the pairs of left and right singular vectors:

$$\mathbf{p}_k = \mathbf{l}_k, \quad \mathbf{q}_k = \mathbf{r}_k. \tag{14}$$

The total (squared) covariance explained by a single pair of patterns is σ_k^2, so that the squared covariance fraction, or the percentage of the (squared) covariance explained by a pair of patterns, is [from (S6)]:

$$\mathrm{SCF}_k \equiv \frac{\langle a_k(t), b_k(t)\rangle^2}{\|C_{sz}\|_F^2} = \frac{\sigma_k^2}{\sum_{l=1}^{r}\sigma_l^2}. \tag{15a}$$

Similarly, one can show that the cumulative squared covariance fraction of C_{sz} explained by the leading N modes is

$$\mathrm{CSCF}_N = \frac{\sum_{k=1}^{N}\sigma_k^2}{\sum_{k=1}^{r}\sigma_k^2}. \tag{15b}$$

Stewart (1973, Theorem 6.7, p. 322) showed that for a given matrix C_{sz}, the matrix $\tilde{C}$ of rank N that minimizes $\|C_{sz} - \tilde{C}\|_F$, and hence the SCSF, is given by the first N terms in the SVD of C_{sz}. For any of the methods discussed, the synthetic covariance matrix derived using N modes has rank N. Hence, SVD is an "optimal" method in that it explains the maximum possible CSCF with N modes.

The CSCF is an analog for a covariance matrix of the "cumulative variance fraction" explained by the leading modes in PCA analysis. If the left and right fields s and z are identical, (S5) implies that PCA and SVD will yield equivalent results with the singular values σ_k equal to the mode variances λ_k (the eigenvalues of the variance matrix). Note that in this case, CSCF, as obtained from (15b), approaches unity more rapidly with the incorporation of additional modes than the cumulative variance fraction $\mathrm{CVF}_N = \sum_{k=1}^{N}\lambda_k / \sum_{k=1}^{r}\lambda_k$, because the variances are *squared* in the CSCF, emphasizing the leading modes with larger mode variances. Thus, large CSCF is less significant than large CVF.

Once $a_k(t)$ and $b_k(t)$ are obtained from SVD, we can generate heterogeneous and homogeneous correlation maps, as defined in section 2. If the time series have been normalized, the left heterogeneous correlation map is proportional to the left singular vector. From (1d) and (S1):

$$r[\mathbf{s}(t), b_k(t)] = \langle \mathbf{s}(t)b_k(t)\rangle / \langle b_k^2(t)\rangle^{1/2}, \quad \text{if } \langle s_i(t)^2\rangle = 1 \quad \forall i$$

$$= C_{sz}\mathbf{q}_k/\langle b_k^2(t)\rangle^{1/2},$$
$$= (\sigma_k/\langle b_k^2(t)\rangle^{1/2})\mathbf{p}_k. \quad (16a)$$

Note that the kth left homogeneous correlation map is *not* proportional to $\mathbf{p}_k$; in fact, from (1c) one can deduce that for a normalized time series

$$r[\mathbf{s}(t), a_k(t)] = C_{ss}\mathbf{p}_k/\langle a_k^2(t)\rangle. \quad (16b)$$

The schematic flow chart in Fig. 1 illustrates the steps involved in SVD. In this figure, a vector is represented by a typical component of that vector with subscript i for the left field and subscript j for the right field. First we compute the SVD of the covariance matrix between the two fields (triangle). The gridpoint data are projected onto its singular vectors to obtain the expansion coefficients $a_k(t)$ and $b_k(t)$. Last, the expansion coefficients can be correlated with the gridpoint data to find the correlation maps.

b. *Canonical correlation analysis*

Canonical correlation analysis is a technique that isolates the linear combination of data from the left field and the linear combination of data from the right field that have the maximum correlation coefficient. This pair of time series is more strongly correlated than the expansion coefficients of the leading pair of patterns deduced from SVD, but explains a smaller fraction of the covariance between the two fields.

To identify the linear combinations $a(t)$ and $b(t)$ that are maximally correlated, we express them in terms of unknown weight vectors multiplying time series of the left and right fields:

$$a(t) = \mathbf{u}^T\mathbf{s}(t), \quad (17a)$$
$$b(t) = \mathbf{v}^T\mathbf{z}(t). \quad (17b)$$

Without changing the spatial structure of the left and right weight vectors (and hence the correlation between the left and right expansion coefficients), we can require that $a(t)$ and $b(t)$ be normalized to have variance 1 (because the correlation coefficient is unchanged when the expansion coefficients are multiplied by constants). To retain the terminology introduced in section 2, $a(t)$ and $b(t)$ will be called the left and right expansion coefficients, respectively (rather than the predictor/predictand terminology used in many papers on this subject), and $\mathbf{u}$ and $\mathbf{v}$ the left and right weight vectors (rather than the commonly used term "canonical vectors").

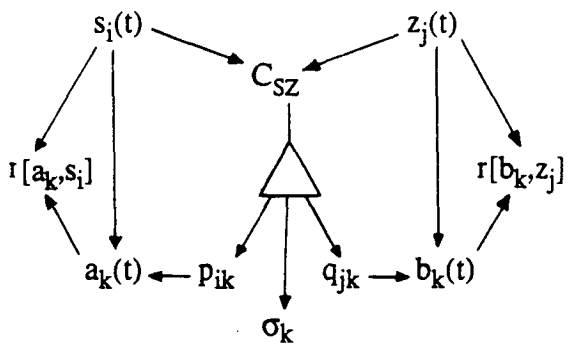

FIG. 1. A schematic outline of the computational procedures used in direct SVD. The SVD operation itself is represented by the triangle. All vectors are written in component form.

In SVD, the weight vectors are constrained to be mutually spatially orthogonal, so generalized Fourier analysis guarantees that the gridpoint data can be recovered as the sum of the expansion coefficients multiplied by their respective weight vectors; that is, the patterns are the same as the weight vectors. In CCA the expansion coefficients are constrained to be *temporally* uncorrelated, but in general the weight vectors are *not* spatially orthogonal. Hence, the patterns are *not* the same as the weight vectors. The weight vectors indicate which gridpoint values are dominant in forming the expansion coefficient. As will be shown in section 3b.(2), for CCA the patterns are the homogeneous correlation maps, which show how individual gridpoint values are correlated with the expansion coefficient. As with SVD, the left and right weight vectors and homogeneous correlation maps are important pairs of fields to examine when using CCA.

There are two mathematical approaches to CCA—a "classical" treatment found in many texts on multivariate analysis (e.g., Anderson 1958, among others) and an approach due to Barnett and Preisendorfer (1987) in which the time series of each field is filtered by projection onto a leading subset of its EOFs and then the maximum correlation between linear combinations of the filtered time series of the two fields is sought.

1) THE CLASSICAL APPROACH TO CCA

In the classical approach, the weight vectors are found by solving the following constrained maximization problem using Lagrange multipliers:

$$\mathbf{r}[a(t), b(t)] = \langle a(t)b(t)\rangle$$
$$= \mathbf{u}^T\langle \mathbf{s}(t)\mathbf{z}^T(t)\rangle\mathbf{v} = \max,$$

with

$$\langle a^2(t)\rangle = \mathbf{u}^T\langle \mathbf{s}(t)\mathbf{s}^T(t)\rangle\mathbf{u} = 1 \quad \text{and}$$
$$\langle b^2(t)\rangle = \mathbf{v}^T\langle \mathbf{z}(t)\mathbf{z}^T(t)\rangle\mathbf{v} = 1. \quad (18)$$

We can simplify the appearance of the constraints by defining "normalized" weight vectors $\hat{\mathbf{u}}_k$ and $\hat{\mathbf{v}}_k$ such that

$$\mathbf{u}_k = C_{ss}^{-1/2}\hat{\mathbf{u}}_k, \quad (19a)$$
$$\mathbf{v}_k = C_{zz}^{-1/2}\hat{\mathbf{v}}_k. \quad (19b)$$

Substituting (19) into the maximization problem (18)

transforms it into the equivalent maximization problem:

$$\langle a(t)b(t)\rangle = \hat{\mathbf{u}}^T\hat{A}\hat{\mathbf{v}} = \max, \quad \hat{A} = C_{ss}^{-1/2}C_{sz}C_{zz}^{-1/2} \tag{20a}$$

with

$$\langle a^2(t)\rangle = \hat{\mathbf{u}}^T\hat{\mathbf{u}} = 1 \quad \text{and} \quad \langle b^2(t)\rangle = \hat{\mathbf{v}}^T\hat{\mathbf{v}} = 1, \tag{20b}$$

which has a form identical to the maximization problem (5). Therefore, the solution is given by the SVD of $\hat{A}$, as pointed out by Nicholls (1987) and as indicated in Table 1. The normalized weight vectors are the singular vectors of $\hat{A}$; the weight vectors are found from (19).

The singular values ρ_k are usually referred to as the *canonical correlations.* They are the correlation coefficients between the left and right expansion coefficients. The left and right expansion coefficients $a_k(t)$ and $b_k(t)$ are obtained by substituting the left and right weight vectors $\mathbf{u}_k$ and $\mathbf{v}_k$ into (17). They obey the temporal orthogonality relations shown in Table 1. Each expansion coefficient pair has the maximum possible correlation coefficient subject to the restriction that the left expansion coefficient is uncorrelated with all the previous left expansion coefficients, and similarly for the right expansion coefficient.

The patterns are chosen such that the rms difference between the real and the synthetic time series for both left and right fields at each grid point is minimum. Note that this is a *different* and perhaps more natural fitting criterion than used for the SVD method. Mathematically, we choose p_{ik} to minimize $\langle[s_i(t) - \tilde{s}_i(t)]^2\rangle$; similarly for the q_{jk} using z. The temporal orthonormality of the expansion coefficients makes this easy; the minimum error is obtained by choosing the patterns to be the temporal projection of the time series onto the expansion coefficients:

$$p_{ik} = \langle s_i(t)a_k(t)\rangle \quad \text{or} \quad \mathbf{p}_k = \langle \mathbf{s}(t)a_k(t)\rangle, \tag{21a}$$

$$q_{jk} = \langle z_j(t)b_k(t)\rangle \quad \text{or} \quad \mathbf{q}_k = \langle \mathbf{z}(t)b_k(t)\rangle. \tag{21b}$$

Since the expansion coefficients have unit variance, *the homogeneous correlation maps are the same as the patterns for normalized data,* as indicated in Table 1. Expressing the expansion coefficients as inner products of the fields and the weight vectors using (2) in (21) implies that

$$\mathbf{p}_k = C_{ss}\mathbf{u}_k, \tag{22a}$$

$$\mathbf{q}_k = C_{zz}\mathbf{v}_k. \tag{22b}$$

In geophysical applications, the number of observation times may be less than the number of grid points in the spatial fields. In this case, the matrices C_{ss} and C_{zz} are not of full rank and are not invertible. A solution can still be obtained using a generalized inverse or "pseudoinverse" (see Muller 1982; or Kharti 1976), but this solution is more difficult to calculate and interpret. In fact, unless the number of observation times is much larger than the number of grid points, the same problem shows up in another form (Preisendorfer 1988): the random sampling fluctuations in the covariance matrices create pairs of weight vectors with serendipitous extremely high correlations. An example of this small sample size sensitivity is presented in section 4.

2) CCA IN THE BASIS OF PRINCIPAL COMPONENTS

An alternative approach to CCA (Barnett and Preisendorfer 1987) is preferable because it circumvents the aforementioned problem. The two fields are prefiltered by retaining only the projection of each field on a subset of its EOFs then applying CCA. If all the EOFs of both fields are retained, the results of this method are identical to those of the classical approach to CCA; prefiltering by retaining only a few leading EOFs decreases the number of degrees of freedom in each field and renders the resulting modes more stable with respect to sampling variability. This filtering may introduce systematic biases into the predicted weight vectors; these will be addressed in section 4.

To prefilter the fields, each is expressed in its PC (principal component) basis as defined in section 2. In the PC basis, the projection of the left field onto its N_1 leading EOFs can be written as the N_1-vector $\alpha(t)$ with components $\alpha_1(t), \ldots, \alpha_{N_1}(t)$, and the projection of the right field onto its N_2 leading EOFs is the N_2-vector $\beta(t)$ with components $\beta_1(t), \ldots, \beta_{N_2}(t)$. Note that any linear combination of filtered gridpoint time series can just as well be written in the PC basis as an equivalent linear combination of the expansion coefficients retained. Therefore, CCA can be done in the PC basis using the fields $\alpha(t)$ and $\beta(t)$. The temporal orthonormality of PCA expansion coefficients guarantees that

$$C_{\alpha\alpha} = \langle \alpha(t)\alpha^T(t)\rangle = I \tag{23a}$$

$$C_{\beta\beta} = \langle \beta(t)\beta^T(t)\rangle = I. \tag{23b}$$

Hence, in the PC basis, CCA reduces to finding the SVD of $C_{\alpha\beta}$, and the left and right weight vectors are just the left and right singular vectors $\hat{\mathbf{l}}_k$ and $\hat{\mathbf{r}}_k$ of this matrix. The expansion coefficients are then found as inner products of these weight vectors with the fields (in the PC basis) as indicated in Table 1.

Using (8) the inner product for the left expansion coefficient can be written in the form $\mathbf{u}_k^T\mathbf{s}(t)$, where $\mathbf{u}_k$ can be interpreted as the left weight vector in the gridpoint basis, and is given in Table 1. The expansion coefficients obey the same temporal orthonormality as in classical CCA, since we have just applied to an unconventional pair of fields. Hence, formulas (21) and (22) for the patterns still apply. When applied to the BP left weight vector, (22) yields the left pattern given

in Table 1. Similar formulas are given for the right expansion coefficient, weight vectors, and patterns.

The computational scheme for CCA is outlined in Fig. 2. First a PCA of each field is performed (squares), and then the normalized expansion coefficients $\alpha_m(t)$ and $\beta_n(t)$ of the leading modes are computed. To highlight the fact that only a subset of the PCA modes is retained the squares are darkened. Then the SVD of the covariance matrix between the expansion coefficients is found (triangle), the outputs of which are the canonical correlations $\hat{\rho}_k$ and the weight vectors $\mathbf{u}'_k$ and $\mathbf{v}'_k$ in the PC basis. From these weight vectors, the CCA expansion coefficients $a_k(t)$ and $b_k(t)$ are obtained and correlated with the gridpoint time series to generate the homogeneous correlation maps. Last, the weight vectors $\mathbf{u}_k$ and $\mathbf{v}_k$ in the gridpoint basis can be found from (22).

Had we not normalized the expansion coefficients and truncated the linear combinations before calculating the SVD, we would have recovered precisely the same result as found in SVD. The normalization, which magnifies the role of EOFs that account for little of the variance in their individual fields and are subject to large sampling fluctuations (North et al. 1982), is one way of understanding the poor small-sample stability of CCA. Prefiltering using BP circumvents this difficulty by entirely neglecting the EOFs with small variances. SVD, by maximizing the covariance (rather than the correlation) between the paired linear combinations $[a_k(t), b_k(t)]$ of the two fields, also selects linear combinations of the field variables that not only are well correlated but also each have a large variance and, hence, tend to be dominated by the leading EOFs of the fields. Therefore, it should come as no surprise that the performance of BP-prefiltered CCA and of SVD on the example presented in section 4 are quite similar.

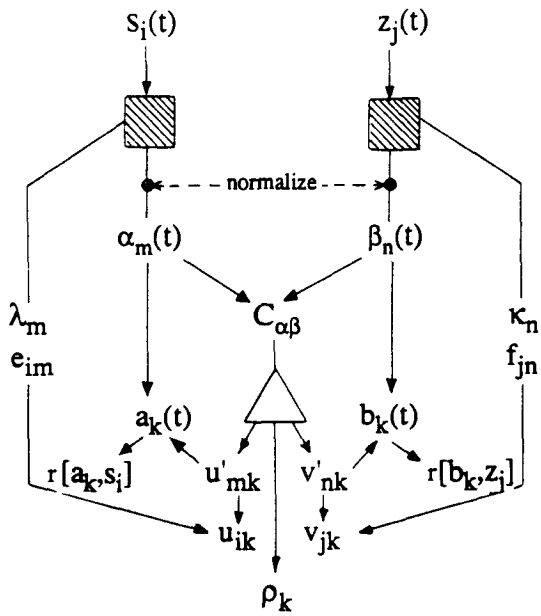

FIG. 2. As in Fig. 1 except for CCA. The squares represent PCA operations. PCA and SVD symbols are darkened for steps in which only a leading subset of the modes in the expansion is retained.

The parameters N_1 and N_2 must be chosen so as to retain only the "important" EOFs. Both N_1 and N_2 must be less than the number T of observation times, or else modes explaining no variance will be retained. In fact, artificially high correlations will likely be obtained unless N_1 and N_2 are both only a small fraction (10% or less) of N_τ, the number of temporal degrees of freedom in the dataset (Davis 1976; also see section 4). Here N_τ is T divided by an integral autocorrelation length that measures how many successive observations tend to be temporally correlated. If successive observations are independent, then the integral autocorrelation length is 1 and N_τ is T. Otherwise, N_τ will be less than T. If too few modes are chosen, correlated parts of the two fields may be discarded (see Joliffe 1982).

In the remainder of this paper, we will choose N_1 and N_2 to be the minimum number of EOFs to account for at least 70% of the variance of their respective fields. This method is referred to as BP70. As discussed in section 4, this leads to weight vectors whose expansion coefficients are both highly correlated and "small-sample stable" (and hence presumably physically meaningful). Tests are performed using the model of section 4, with explained variance thresholds of 70%, 80%, and 90% to determine which threshold gave the optimal trade-off between random errors due to sampling fluctuations (which dominate when the threshold becomes too close to 100% and a large fraction of the EOFs are kept) and systematic errors (which are large when the variance threshold is low and too few EOFs are kept to adequately represent the dominant coupled patterns). It was found for our model problem that the 80% threshold usually led to slightly larger errors than the 70% threshold, while the 90% threshold led to substantially larger errors.

3) AN SVD PREFILTER FOR CCA

In the companion paper (Wallace et al. 1992), SVD is compared with a variation of CCA that uses a straightforward variation of Barnett and Preisendorfer's prefiltering scheme of using only the first few expansion coefficients. In place of the gridpoint data, CCA is applied to time series $\hat{a}_h(t)$ and $\hat{b}_h(t)$ of the expansion coefficients of the N leading SVD modes, but then no further filtering is done. The number N should be chosen to be as small as possible, consistent with the first N modes accounting for most of the covariance between the fields; in the companion paper the sensitivity of CCA to N. Our Nth CCA solution represents the linear combination of the first N modes of the original SVD solution that maximizes the correlation coeffi-

cients between the respective SST and 500-mb height time series. To recover the weight vectors in the gridpoint basis requires *two* steps. First, we proceed as for classical CCA to find the weight vectors $\hat{\mathbf{u}}_k$ and $\hat{\mathbf{v}}_k$ of length N in the basis of the SVD modes. Second, the $N_s \times N$ matrix P whose columns are the N leading left singular vectors from the SVD analysis and the $N_z \times N$ matrix Q whose columns are the N leading right singular vectors are formed. It follows from (1c, d) and (17) that the weight vectors in the gridpoint basis are

$$\mathbf{u}_k = P\hat{\mathbf{u}}_k, \tag{24a}$$

$$\mathbf{v}_k = Q\hat{\mathbf{v}}_k. \tag{24b}$$

Figure 3 diagrams this computational scheme for CCA. In this diagram, the subscript h is the index of a mode retained after SVD prefiltering, and m and n are indices of the s and z EOFs and expansion coefficients derived from the subsequent PCA. Figure 3 is the same as Fig. 2 except that it is encased by an outer shell in which the SVD operation is performed on the gridpoint fields

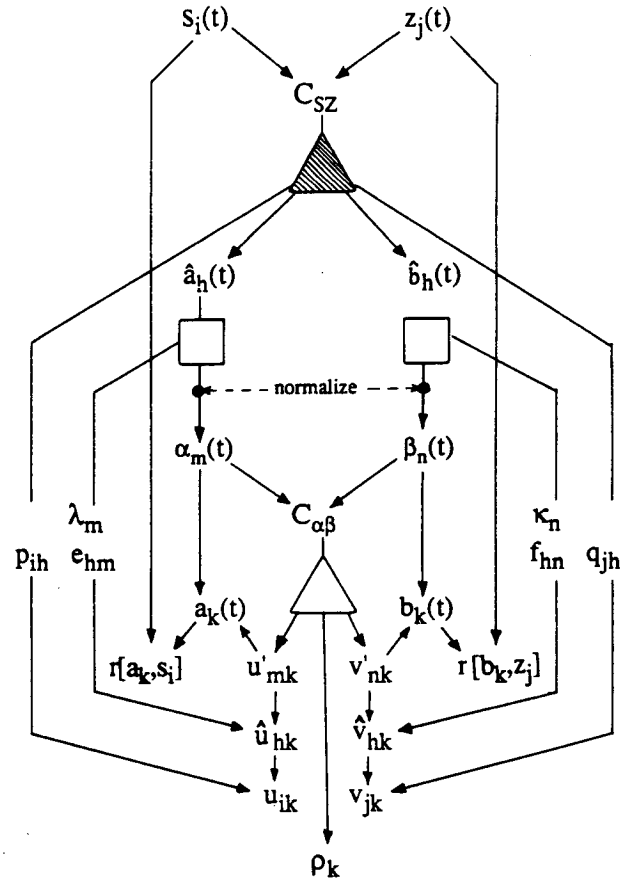

FIG. 3. As in Fig. 2 except for CCA using SVD in place of PCA as a prefilter.

to get prefiltered time series for the first N SVD modes, which are used as input to the CCA. The PCAs are not used to further filter the data, so the squares are not darkened.

c. CPCA and PCA

The PCA separates a field into modes using its own covariance matrix. The terminology of expansion coefficients, weight vectors, and patterns is still applicable to this situation and leads transparently to all of the formulas given in Table 1 for CPCA. In LPCA (one-field PCA based on the EOFs of the left field), the modes of the left field are deduced from PCA, and then patterns in the right field that best correlate with each of these modes are found. Hence, the method is asymmetric. The expansion coefficient used for both fields is the normalized PCA expansion coefficient for the left field. The left weight vector and pattern follow directly from PCA of the left field. Since the expansion coefficient does not depend on the right field, the right weight vector is zero. The right pattern is again found from (21) and (8) since the expansion coefficients are temporally orthonormal:

$$\mathbf{v}_k = \langle a_k(t)\mathbf{z}(t)\rangle = \lambda_k^{-1/2} C_{sz}^T \hat{\mathbf{u}}_k. \tag{25}$$

The left homogeneous correlation map will be the EOF, while from (25) the right heterogeneous correlation map is the right pattern. Similar results hold for RPCA in which PCA is done on the right field and the expansion coefficients of the resulting EOFs are correlated with the left field.

4. A comparative example of SVD, CCA, BP, CPCA, and OFPCA

The example in this sections illustrates some of the trade-offs intrinsic to each method. No one example is universally applicable, so the conclusions are only a loose guide to the performance of the methods on geophysical datasets.

a. Description of a model coupled system

Let us imagine that time series of SST are measured at two "rakes" of observing stations (Fig. 4). Each rake consists of a series of N stations, all at the same longitude but at different, fairly closely spaced latitudes. Such a rake might be set up to detect equatorially trapped wave modes in the atmosphere or ocean. The two rakes are assumed to be relatively far apart in latitude—one might be in the west Pacific, one in the east Pacific. Suppose that the SST anomaly at each station

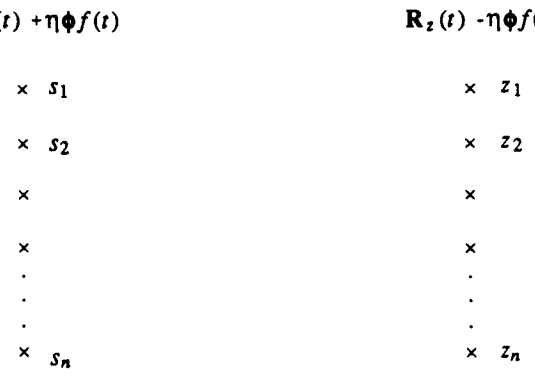

FIG. 4. The model coupled system (explained in the text).

consists of spatially red noise (which is correlated with neighboring stations but not with stations in the other rake) plus a sinusoidal oscillation (the "signal") in time whose amplitude depends on latitude and whose phase is anticorrelated between the rakes. For instance, imagine that when the west Pacific is warmer, the east Pacific is colder, with maximum temperature anomalies occurring on the equator with a Gaussian decline of the anomalies with latitude. Now compare how well the various methods can isolate the spatial pattern of SST anomalies in the two rakes associated with the signal.

We generate synthetic time series for the two rakes, assuming that each time the red noise is completely uncorrelated with the previous observation time. The time series have the form

$$s(t) = \mathbf{R}_s(t) + \eta\phi f(t), \tag{26a}$$

$$z(t) = \mathbf{R}_z(t) - \eta\phi f(t). \tag{26b}$$

Here $\mathbf{R}_s(t)$ and $\mathbf{R}_z(t)$ are vectors of temporally uncorrelated but spatially red noise of unit variance. The latitudinal variation of the signal on each rake is given by the vector ϕ, which is normalized so that its rms component amplitude is one. The signal-to-noise ratio is η. It is assumed that $f(t)$ is a periodic function of t with mean zero and rms value 1 and that the total length of time for which observations have been taken is an integral multiple (or at least much longer than) of the signal period so that $\langle f^2(t)\rangle = 1$. For definiteness,

$$f(t) = 2^{1/2}\sin(\omega t); \tag{27}$$

any other similarly normalized periodic signal would give essentially identical results.

We will compare the methods applied to three different signal shapes. In all of these cases, a rake with more stations is assumed to sample the same signal (and the same red noise) at an increased spatial resolution. The *bell* signal is an off-center Gaussian:

$$\phi_i = c_b \exp(-y_i^2/2)$$

$$y_i = 5(i-1)/(N-1) - 1, \quad i = 1, \cdots, N, \tag{28a}$$

where c_b is a normalization constant. The *half-wave* signal is a half-wavelength cosine wave over the span of the rake,

$$\phi_i = c_h \cos(\pi y_i/2). \tag{28b}$$

The *spike* signal is an off-center spike at a single station,

$$\phi_i = \begin{cases} N^{-1/2}, & y_i = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{28c}$$

To generate the red noise vector $\mathbf{R}(t)$ for either field, at each time t_n an independent pseudorandom Gaussian white noise vector $\mathbf{W}(t_n)$ is generated. Then we use the Markov process:

$$R_1(t_n) = W_1(t_n)$$

$$R_i(t_n) = aR_{i-1}(t_n) + (1-a^2)^{1/2} W_i(t_n). \tag{29}$$

The red noise generated has mean zero, variance one, and

$$\langle \mathbf{R}(t)\mathbf{R}^{\mathrm{T}}(t)\rangle = V(a), \quad V_{ij}(a) = a^{|i-j|}. \tag{30}$$

If the number of stations is varied (corresponding to a shorter distance between stations) we vary a so as to maintain a fixed *noise redness length* L (the separation at which the red noise correlation drops to e^{-1}). If length is measured in signal half-widths, the rake is 5 units long and the interstation spacing is $5/(N-1)$, so

$$a = \exp\{-5/[(N-1)]L\}. \tag{31}$$

b. *Method comparison for an infinite time record*

In this subsection we investigate how the methods perform if the number T of observation times is infinite, so that the covariance matrices estimated from the data are the true covariance matrices. In the next subsection, we look at the geophysically relevant case in which T is finite, so that the covariance matrices estimated from the data differ from the true covariance matrices as a result of sampling fluctuations.

It is not entirely obvious how to interpret the output of some of these methods in terms of correlated signals in the two fields. Probably the most appropriate output to look at is the left and right patterns. By construction, the first left pattern when multiplied by the first left expansion coefficient gives an estimate of the left field that is optimal (in a sense that depends on the method but generally maximizes some measure of coupling between the fields over all estimates based on a single pattern); it is similar for the right pattern. The coupling

between the fields is entirely in the signal, so the patterns should reflect the signal. For the PCA-based methods, the patterns also incorporate aspects of the red noise within each rake as well as the signal. Since the noise characteristics are unknown, we still interpret the correlated signals in the left and right fields to be proportional to the left and right patterns. For comparison, all of the patterns are normalized to have mean-square component amplitude one. The different types of normalization used for different methods preclude a direct comparison of the signal amplitude predicted by the different methods. For an infinite time record, SVD and CCA can detect the signal exactly in both fields, while the other methods have systematic prediction biases because they incorporate red noise characteristics into the patterns. This can be demonstrated by calculating the exact covariance matrices from

$$C_{ss} = C_{zz} = V(a) + \eta^2 \phi\phi^{\mathrm{T}} \tag{32a}$$

$$C_{sz} = -\eta^2 \phi\phi^{\mathrm{T}}. \tag{32b}$$

All of the methods can be carried out knowing these covariance matrices use the algorithms in Table 1. In Fig. 5, results are shown for rakes with $N = 36$ points each, $\eta = 1$, and a noise redness length $L = 1$. The normalized patterns for SVD and CCA are identical to the signal. BP70 exhibits slight biases due to the prefiltering of the two fields. For CPCA both patterns are biased, for LPCA the left pattern is biased but the

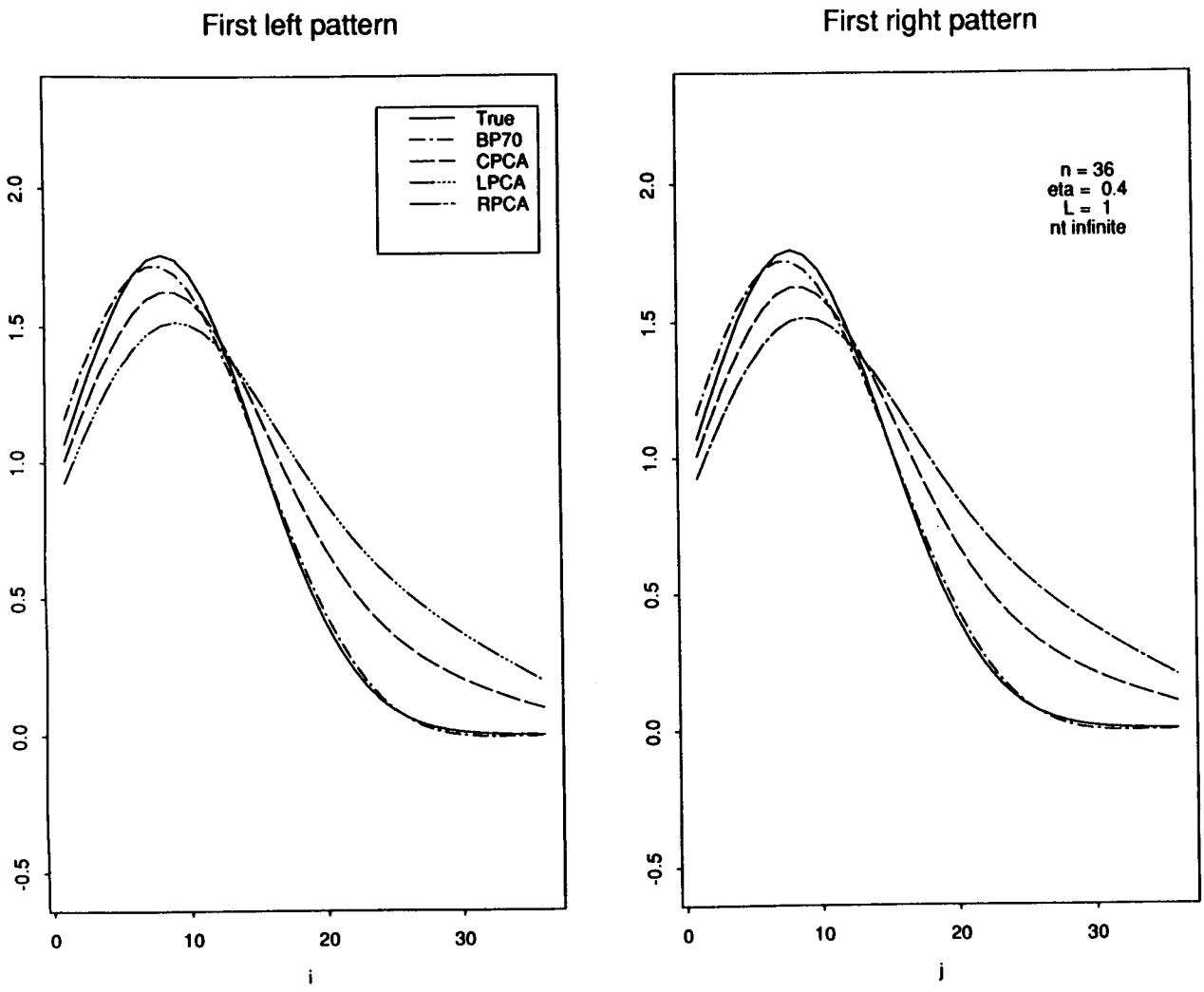

FIG. 5. The true (exact) bell signal in the left and right fields and the normalized patterns for $N = 36$, $\eta = 1$ for BP70, CPCA, LPCA, and RPCA given an infinite length time series. The SVD and CCA patterns are identical to the exact signal.

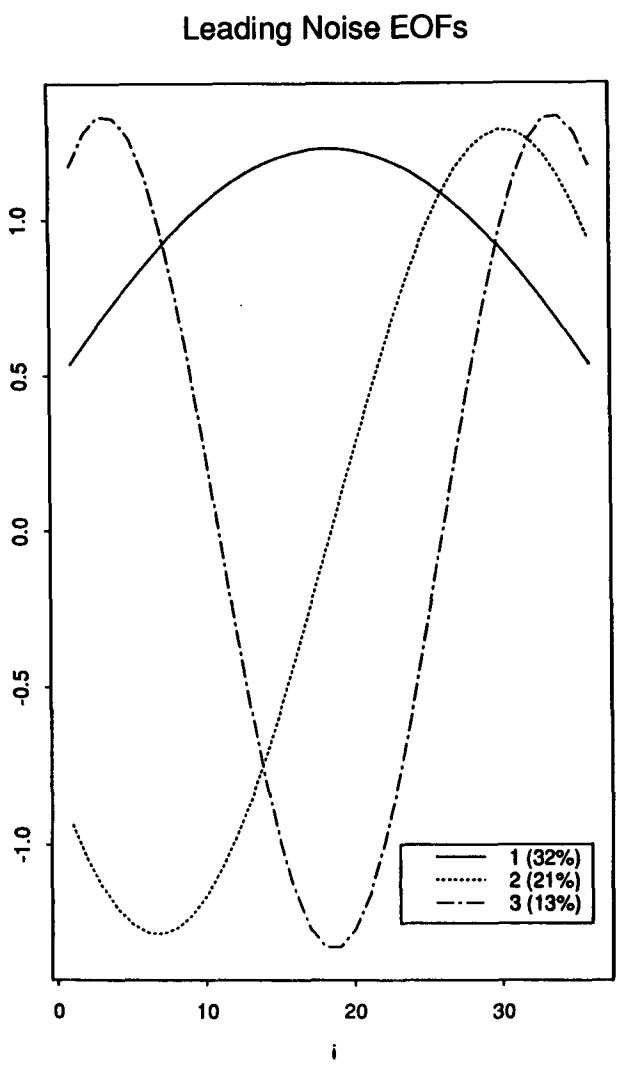

FIG. 6. The leading three EOFs of the noise covariance matrix.

right pattern is correct, and for RPCA the right pattern is biased but the left pattern is correct. These biases can be understood as follows. The CPCA patterns maximize the explained variance over both fields combined. If the noise is highly spatially red, most of the noise variance can be explained by a few modes (so that the patterns are biased toward these modes). Figure 6 shows the first, second, and third EOF of the pure noise variance matrix $V(a)$ for each rake. These three EOFs explain 32%, 21%, and 13% of the variance, respectively. One can see that the CPCA patterns are biased toward the first noise EOF. Similarly, the left patterns in LPCA are just the EOFs of the left field, which reflect the leading modes of the left noise variance matrix $V(a)$ in addition to the signal. The right patterns in LPCA are exact, because the only information used about the right field is its cross correlation matrix with the left field. Since the only correlation between the fields is through the signal, and the left pattern, though biased, is correlated with the signal, the right pattern will just be the signal. Analogous results apply to RPCA.

A corollary of the above discussion is that the leading patterns for CPCA, LPCA, and RPCA may in fact have almost no relation to the signal, especially if the signal is weak. For instance, suppose that LPCA is performed when the signal is orthogonal to the leading EOF of the noise variance and it is projected principally onto the second noise EOF. Now consider the full left covariance matrix C_{ss} of signal plus noise. For a small signal-to-noise ratio, the leading EOF will be approx-

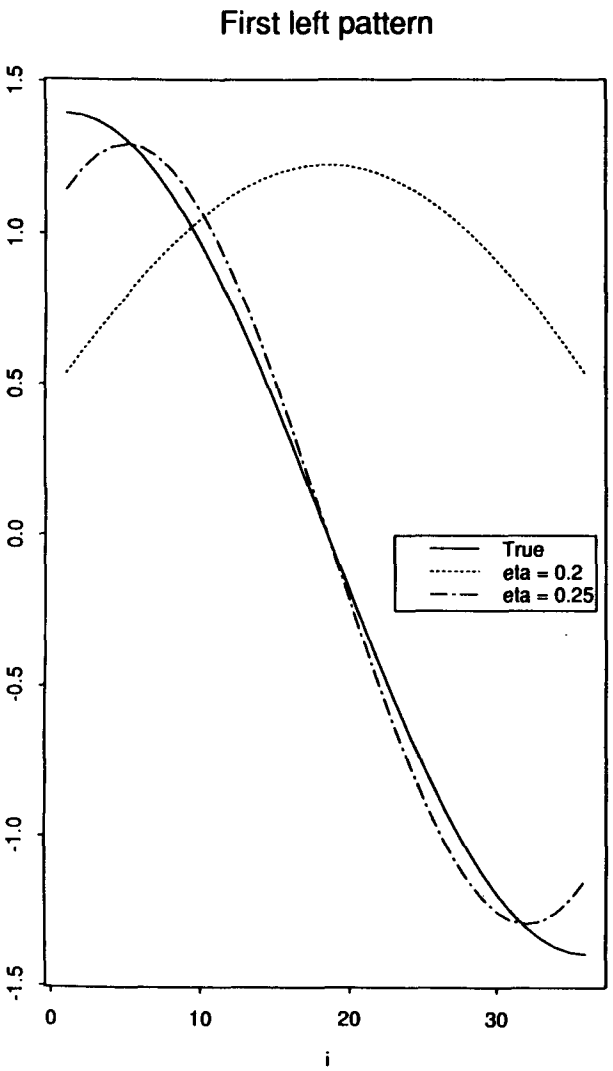

FIG. 7. The half-wave signal and the "exact" patterns for CPCA as functions of the signal-to-noise ratio.

imately the leading EOF of the noise alone, which can still explain a plurality of the variance. Hence, the left pattern will have no resemblance (and in fact will be orthogonal to) the signal. For a larger signal-to-noise ratio, the combination of the signal and the second noise EOF will have larger variance and the pattern will better resemble the signal. CPCA also can be "fooled" in this way, but only at a lower signal-to-noise ratio, because CPCA also uses the cross covariances between the fields, which are entirely due to the signal.

To illustrate this point, the left pattern for CPCA and LPCA is shown in Fig. 7 for the "half-wave" signal at $\eta = 0.2$ and 0.25 with $N = 36$ and noise redness length $L = 1$ as above. For $\eta = 0.25$, the pattern is quite close to the true signal, but for $\eta = 0.2$, the pattern is entirely different—in fact, it is the leading EOF of the noise as seen in Fig. 6. The bias in CPCA, LPCA, and RPCA decreases rapidly as η is increased (see the lower curves on Fig. 9). For the "spike" signal, which has projections of approximately equal magnitude on all of the noise EOFs, no method was fooled but there was a strong bias in the leading patterns toward the nearly uniform leading noise EOF at small signal-to-noise ratios.

c. *Method comparison for a finite time record*

In geophysical applications, the number T of observation times may not greatly exceed the number of locations N at which each field is measured. The sampling fluctuations in the covariance matrices can greatly reduce the skill of all of these methods at isolating the signal. Using synthetically generated red noise, each of

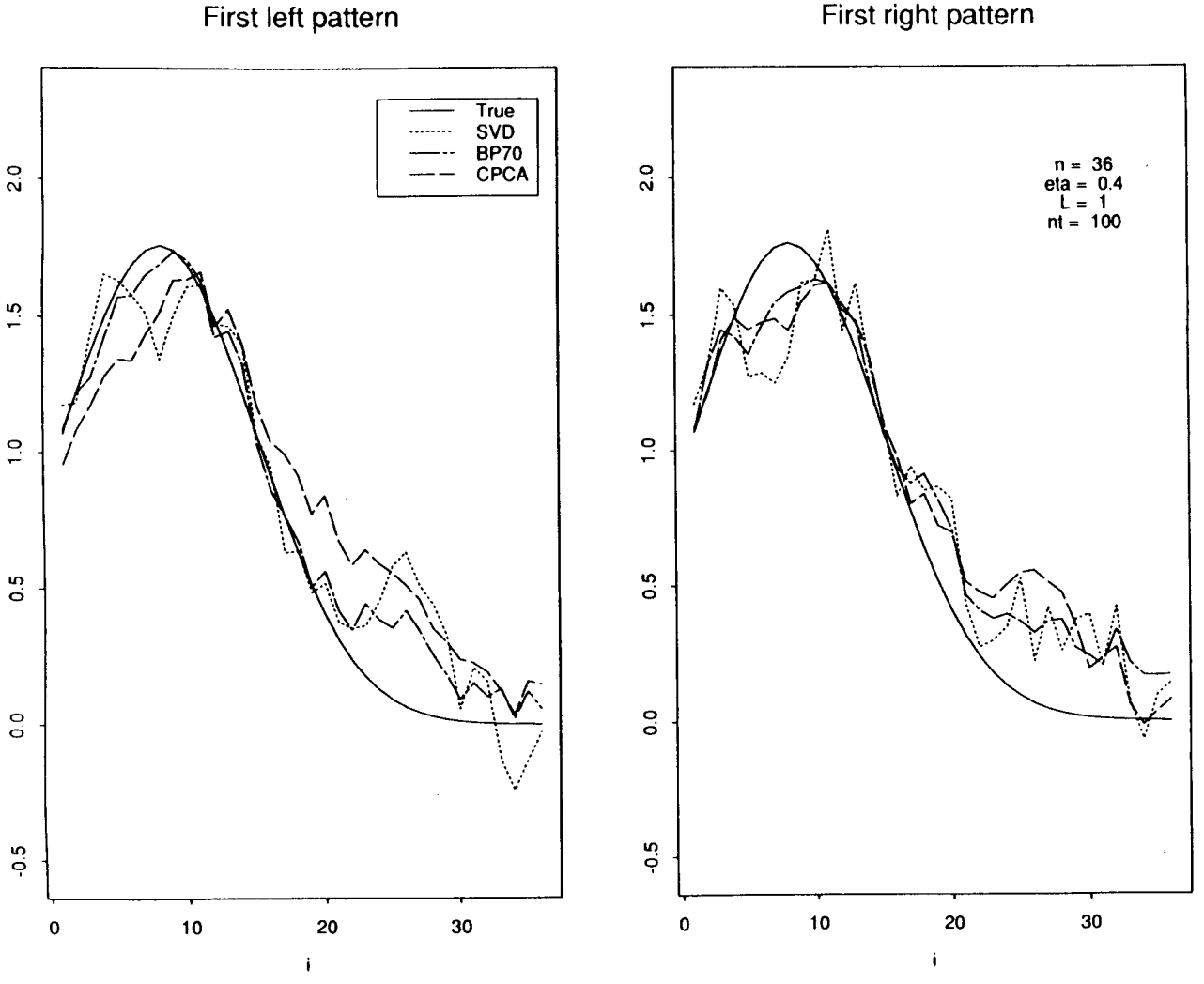

FIG. 8. The left SVD, BP70, and CPCA patterns for an pseudorandom time series of 100 independent observations: $N = 36$, $\eta = 0.4$.

these methods is applied to time series of finite length. An ensemble of 20 such realizations was used to determine the "error" ϵ for each method, defined as the square root of the average over the 20 realizations of the mean-square difference of the components of the normalized left and right patterns from the signal. It was found (at least when the signal-to-noise ratio was not too small) that the average over all of the realizations of the patterns was very close to the infinite time record patterns. Consequently, it is useful to decompose the finite time record errors for each method into a "random" component ϵ_r (the rms difference of the pattern from the $T = \infty$ pattern) and a systematic component ϵ_s due to the method bias. The sum of the square of these errors would be exactly equal to ϵ^2 if the average patterns were the $T = \infty$ patterns; it is slightly larger when this is not exactly the case.

Figure 8 shows the SVD, CCA, BP70, and CPCA left patterns derived for one typical realization with the bell signal, $N = 36$, $\eta = 1$, and $T = 100$. The right patterns show the same behavior. The random error in all the methods is noticeable. Figure 9 shows the rms and systematic errors for 25 realizations of the bell signal with the same rake as the signal-to-noise ratio is varied. As η is decreased, both the total and the systematic errors increase quite rapidly. If instead the number of observations T is varied (Fig. 10) with $\eta = 0.4$, the systematic component of the error begins to dominate the CPCA, LPCA, and RPCA errors for T on the order of $10N$ or larger, and SVD and BP70 give superior results. For $T = 800$, CCA also gives good results. Of all methods, CPCA consistently has the smallest random errors (as little as half that of other methods) because it blends together all of the data and

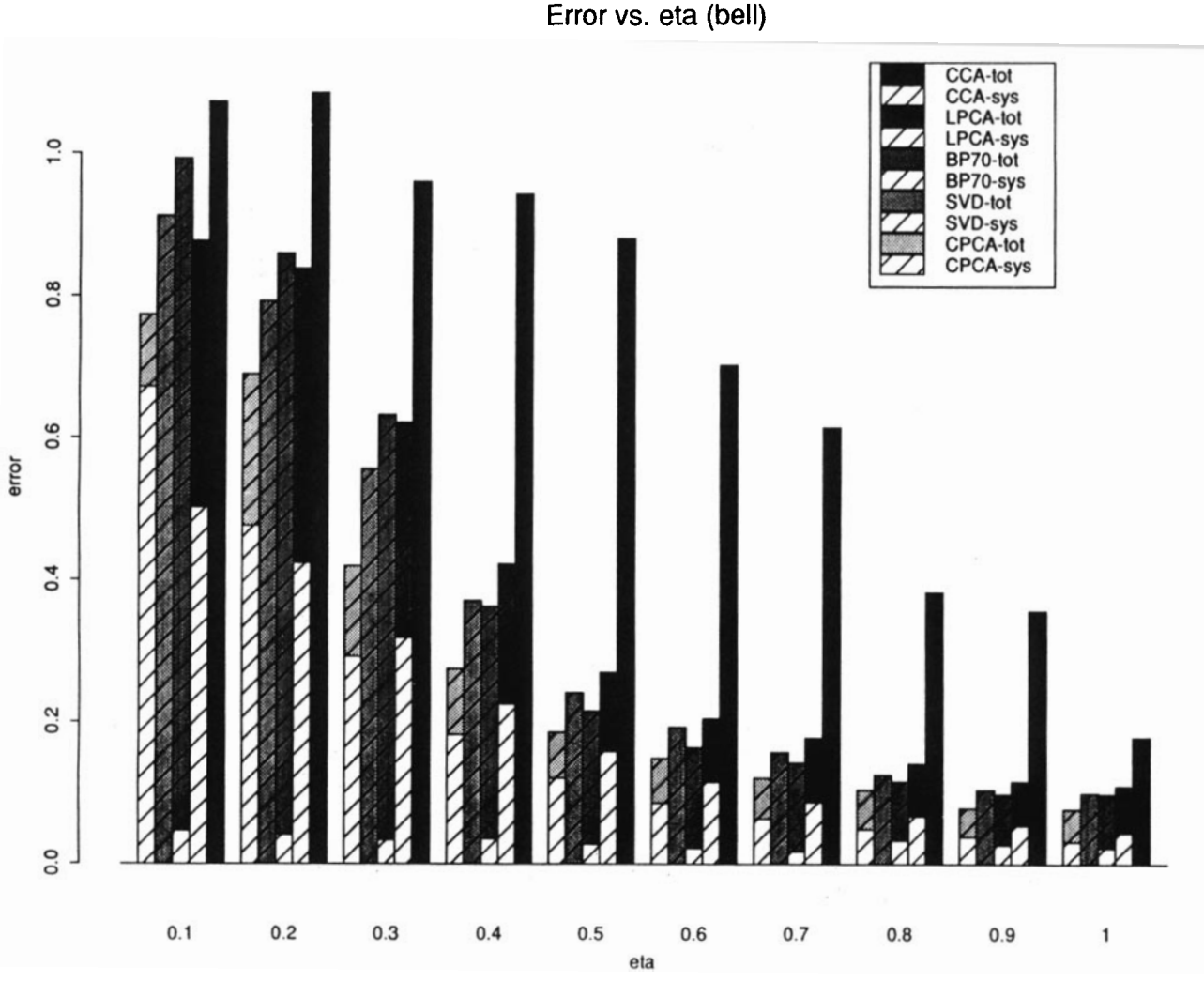

FIG. 9. Rms error ϵ in detecting the bell signal using ensembles of 25 realizations as a function of η, with $T = 100$, $N = 36$.

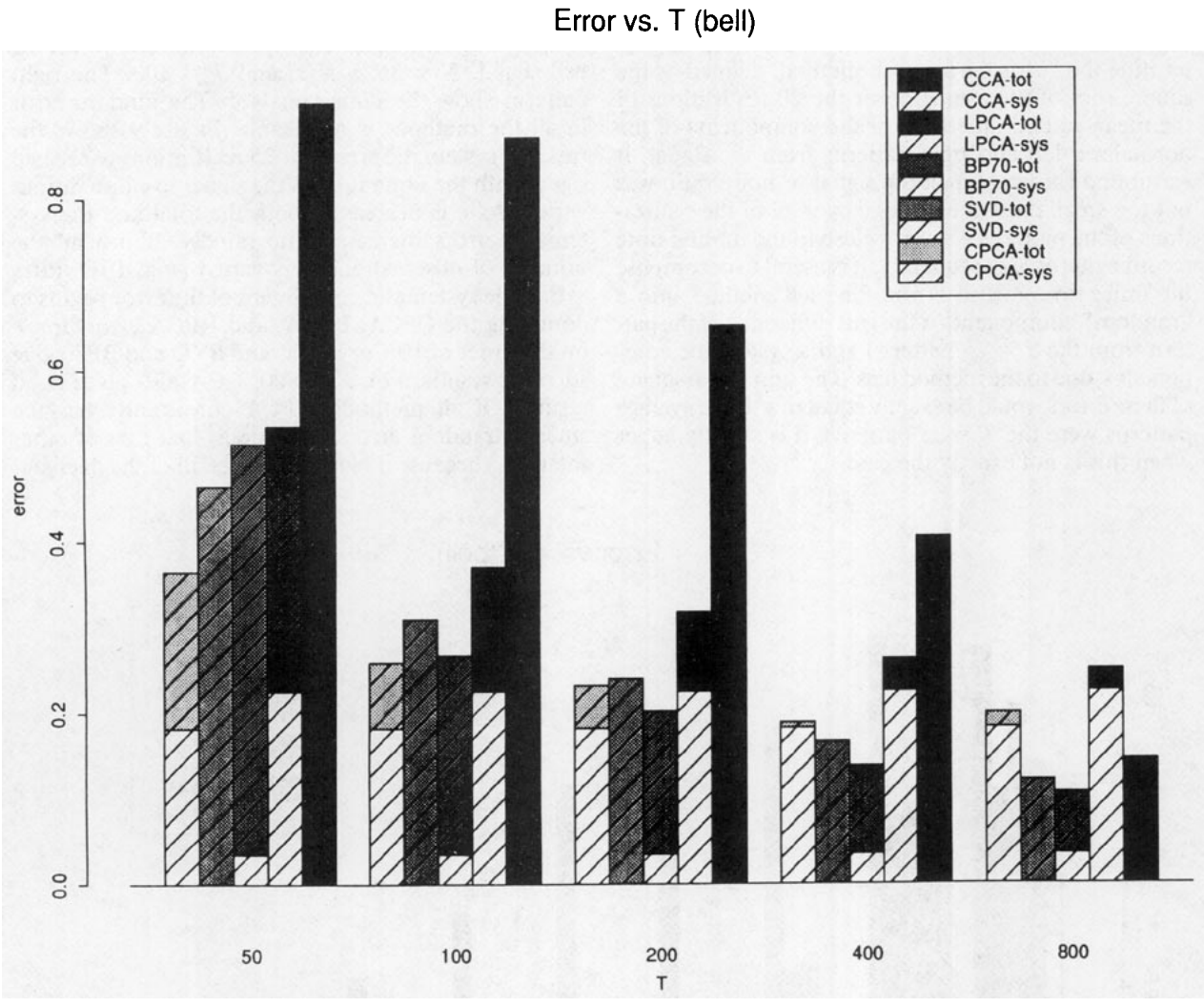

FIG. 10. As in Fig. 9 except as a function of T with $\eta = 0.4$, $N = 36$.

hence has four times as many covariances in which to find a signal. Classical CCA, on the other hand, is extremely prone to random errors with T/N on the order of 10 or less, as it is here. BP70 and SVD have comparable total errors, which are slightly less than those of LPCA in this example. Essentially this is because all of them weight modes that dominate the covariance, which is comparatively small-sample stable. LPCA and RPCA use just the leading EOF of one of the fields, which is the most small-sample stable of all EOFs, but then are vulnerable to the full sampling fluctuations of C_{sz}. SVD finds the coupled modes that best explain the cross covariance between the fields. These modes must explain enough of the variance in both fields individually to also be relatively small-sample stable. The BP method just truncates the EOFs to the few dominant (hence, most small-sample stable) EOFs in both fields before any CCA analysis is done, so it is also a small-sample stable method. The LPCA, RPCA, and (to a somewhat lesser extent) CPCA may have significant systematic biases, while SVD and CCA have no systematic bias (Fig. 9). The systematic bias of BP70 is generally much smaller than the random errors. In general, BP70 and SVD give quite similar patterns in the individual realizations.

With the normalizations used here, the errors are very insensitive to rake size N. With fixed $T = 100$ and $\eta = 0.4$, we compared the rms errors over 25 realizations for $N = 11$, 21, 36, 56, and 81. The results (Fig. 11) show essentially no dependence of the total or systematic errors on N, except for CCA, which had an rms error of 0.5 (somewhat worse than other methods) for $N = 11$ ($T/N = 9$) but had an error of one, corresponding to no skill in finding the coupled signal

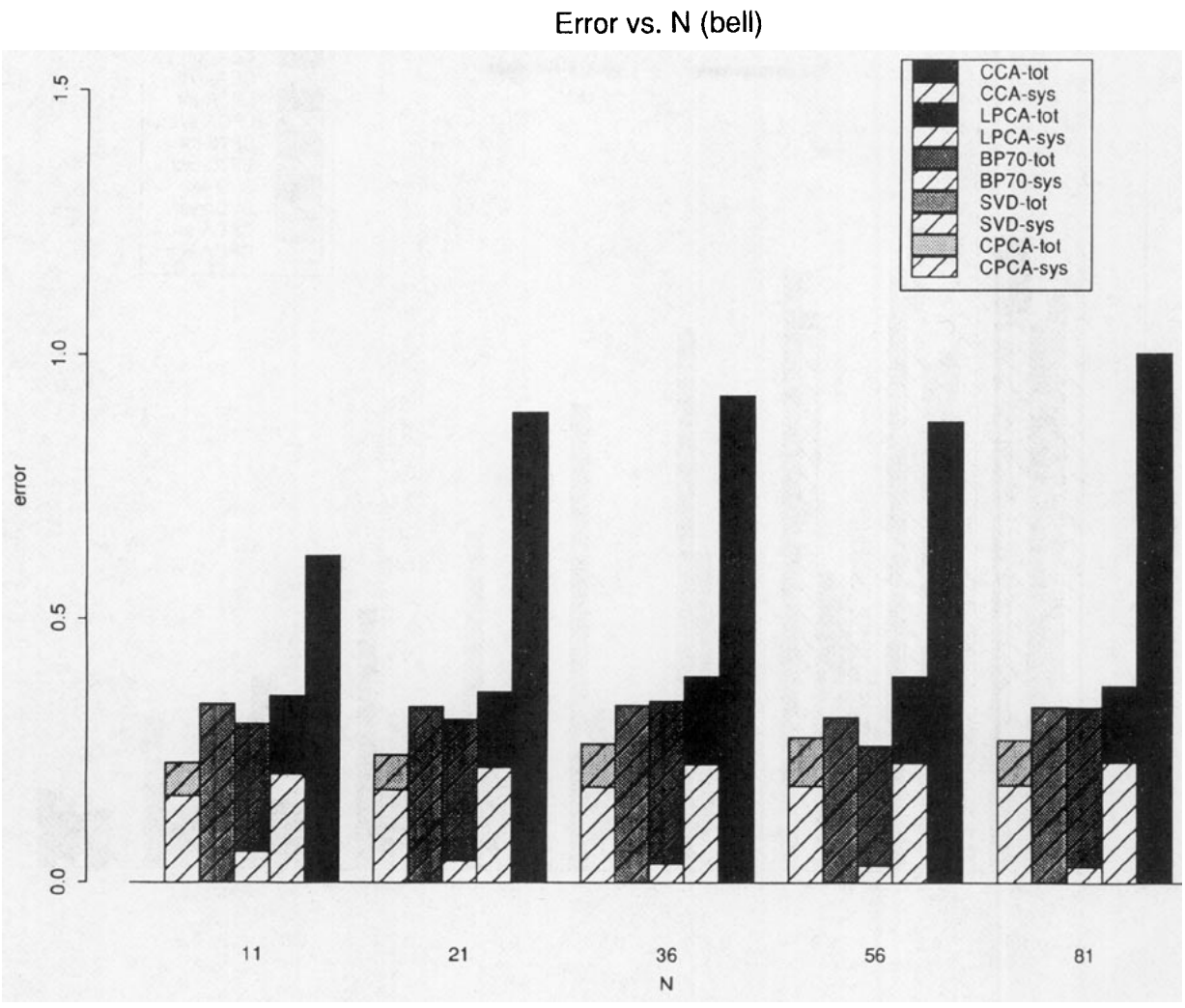

FIG. 11. As in Fig. 9 except as a function of N with $\eta = 0.4$, $T = 100$.

for $N = 36$ ($T/N = 3$) or more, in accordance with Fig. 10.

The half-wave signal produces generally similar results to the bell signal for $\eta > 0.3$ (Fig. 12). For smaller signal-to-noise ratios, SFPCA and CPCA frequently err because they do not choose the signal-containing mode. However, the random sampling fluctuations cause comparable errors in SVD and BP70. CCA is again uncompetitive.

For the spike signal (Fig. 13), CCA does the best at all signal-to-noise ratios. This is not surprising. The spike projects significantly onto all of the EOFs of the noise field, so methods that favor the dominant modes filter out much of the correlation between the fields along with the noise. However, for $\eta \geq 0.4$, SVD and BP70 do not have significantly larger errors than CCA, and for $\eta \geq 0.6$ CPCA also is competitive. Note that the CPCA and LPCA errors are largely due to the systematic error, so they would not improve much with further sampling.

5. Concluding remarks

We have reviewed several methods for finding correlated patterns between two fields. Table 1 indicates how all of these methods can be brought into a unified theoretical framework in which it is easy to compare their performance. Each of the methods optimizes a different measure of goodness of fit to the data. The methods were applied to an example in which two rakes of observing stations experience a single spatially varying signal that is perfectly anticorrelated between the rakes, superposed on spatially red noise, which is uncorrelated between the rakes. Results from this example suggest that if the coupled signals are similar to the dominant EOFs of the individual fields, so that the

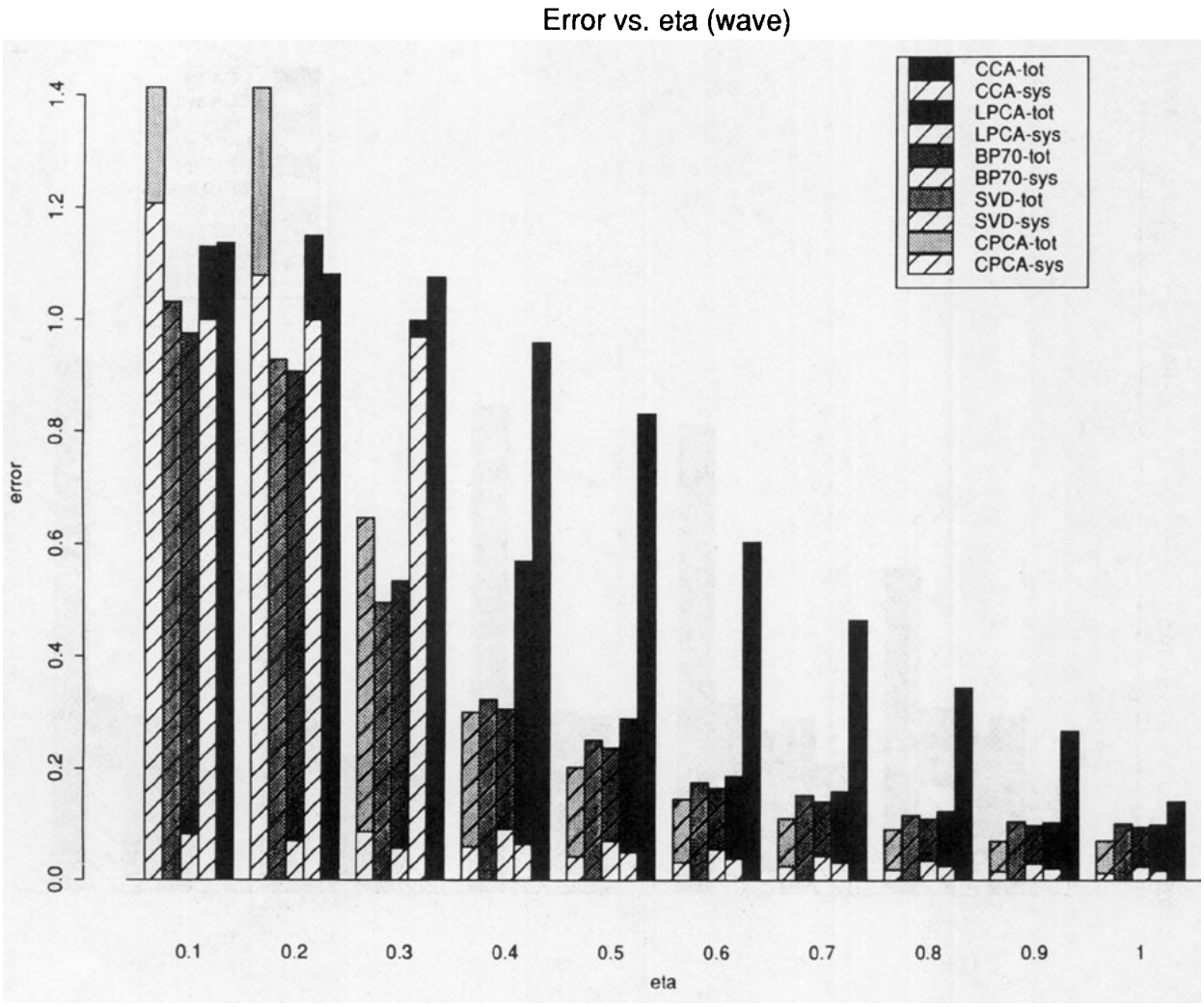

FIG. 12. As in Fig. 9 except using wave signal.

systematic biases in the PCA-based methods are small, then CPCA most accurately isolates the signal since it has the smallest random errors. If the coupled signals are rather dissimilar from the EOFs but are still smooth in the sense that they project principally onto only the leading few eigenmodes of the individual fields, SVD and BP may be superior to CPCA. In fact, in this situation, CPCA may be "fooled" into an entirely misleading first pattern if the signal is almost orthogonal to the leading eigenmodes of the individual fields. SFPCA may also be subject to problems with systematic errors and may exhibit random errors no smaller than found with SVD, therefore SFPCA is uncompetitive. Classical CCA is subject to large random errors, except for spike-like signals that excite all the EOFs of the individual fields equally, in which case it works best of all.

In general, SVD and BP have rather similar error characteristics. SVD is very simple to perform and interpret and requires no user-supplied parameters. The BP method was usually slightly more accurate on our example with a 70% variance threshold but is substantially more complex to implement and requires a criterion (such as the cumulative variance threshold used in this paper) for deciding the number of modes to be kept for each field. It has nonzero but generally insignificant systematic errors. Because of their lack of systematic bias and good general performance, SVD and BP are perhaps the most preferable for general use. A conceptual advantage of both SVD and BP is that, unlike CPCA, they directly produce explicit measures of relatedness between the derived coupled patterns. An important future area of research should be to extend SVD or BP to cases in which one would like to find

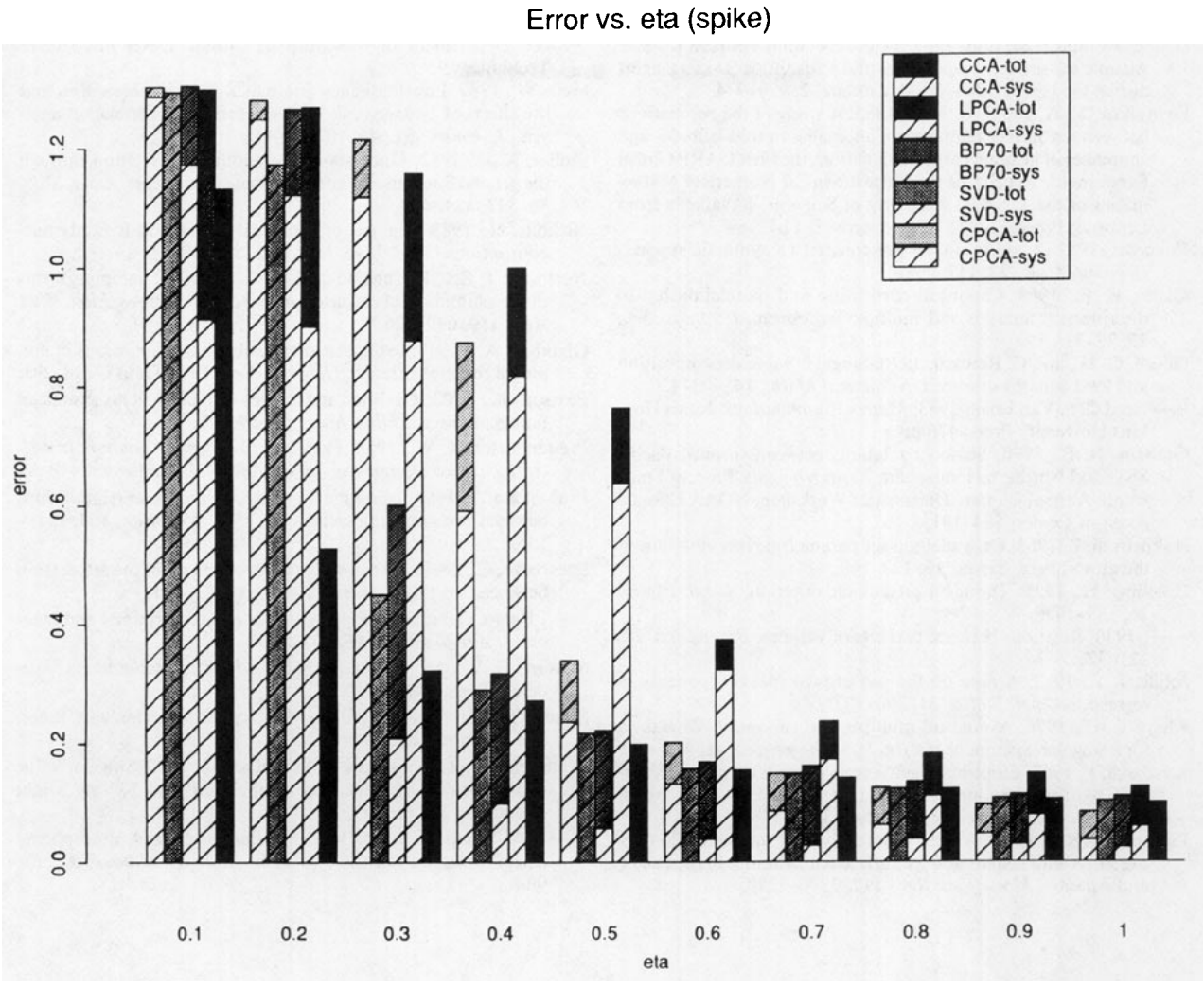

FIG. 13. As in Fig. 9 except using spike signal.

coupled patterns in three or more fields; this type of analysis is straightforward with CPCA.

The companion paper (Wallace et al. 1992) illustrates SVD applied to a geophysical problem, the interannual coupling between wintertime Pacific sea surface temperature anomalies and atmospheric 500-mb height. It compares SVD to CCA and CPCA and PCA, and it shows that SVD clearly isolates the two most important extratropical modes of variability in this case.

Acknowledgments. We would like to thank Dr. Valentin Dymnikov of the Center for Numerical Mathematics of the USSR Academy of Sciences for bringing SVD to our attention. Ms. Yuan Zhang's careful reading helped improve the manuscript. This work was supported by the National Science Foundation through the Climate Dynamics Program under Grant ATM 8822872 and the Meteorology Program under Grant ATM 8858846.

REFERENCES

Anderson, T. W., 1958: *An Introduction to Multivariate Statistical Analysis.* John Wiley & Sons, 374 pp.

Barnett, T. P., 1981: Statistical prediction of Northern American air temperatures from Pacific predictors. *Mon. Wea. Rev.*, **109,** 1021–1041.

——, 1983: Interaction of the monsoon and Pacific trade wind systems at interannual time scales. Part I: The equatorial zone. *Mon. Wea. Rev.*, **111,** 756–773.

——, and R. W. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.*, **115,** 1825–1850.

Cooley, W., and P. Lohnes, 1971: *Multivariate Data Analysis.* John Wiley & Sons, 364 pp.

Davis, R., 1976: Predictability of sea surface temperature and sea-

level pressure anomalies over the Northern Hemisphere. *J. Phys. Oceanogr.*, **6,** 249–266.

Déqué, M., and J. Servain, 1989: Teleconnections between tropical Atlantic sea-surface temperatures and midlatitude 50-kPa heights during the period 1964–86. *J. Climate,* **2,** 929–944.

Dymnikov, V. P., and S. K. Filin, 1985: A study of the correlations between sea-surface temperature anomalies in mid-latitudes and anomalies in heating, based on data from the First GARP Global Experiment. Reprint of the Department of Numerical Mathematics of the U.S.S.R. Academy of Sciences. [Available from Leninskij Prosp., 14; 117901 Moscow B-71.]

Fukuoka, 1951: A study on 10-day forecast (a synthetic report). *Geophys. Mag.*, **22,** 117–208.

Glahn, H. R., 1968: Canonical correlation and its relationship to discriminant analysis and multiple regression. *J. Atmos. Sci.*, **25,** 23–31.

Golub, G. H., and C. Reinsch, 1970: Singular value decomposition and least squares solutions. *Numerical Math.*, **14,** 403–420.

——, and C. F. Van Loan, 1983: *Matrix Computations.* Johns Hopkins University Press, 476 pp.

Graham, N. E., 1990: Seasonal relations between tropical Pacific SSTs and Northern Hemisphere 700-mb heights. Proc. of Fourteenth Annual Climate Diagnostics Workshop. NOAA Climate Analysis Center, 184–191.

Holmstrom, I., 1963: On a method for parametric representation of the atmosphere. *Tellus,* **15,** 127–149.

Hotelling, H., 1935: The most predictable criterion. *J. Ed. Psych.*, **26,** 139–142.

——, 1936: Relations between two sets of variates. *Biometrika,* **28,** 321–377.

Joliffe, I. T., 1982: A note on the use of principal components in regression. *Appl. Statist.* **31,** 300–303.

Kharti C. G., 1976: A note on multiple and canonical correlation for a singular covariance matrix. *Pschychometrika,* **41,** 465–470.

Kutzbach, J., 1967: Empirical eigenvectors of sea-level pressure, surface temperature, and precipitation complexes over North America. *J. Appl. Meteor.*, **6,** 791–802.

Lanzante, J. R., 1984: A rotated eigenanalysis of the correlation between 700-mb heights and sea surface temperatures in the Pacific and Atlantic. *Mon. Wea. Rev.*, **112,** 2270–2280.

Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. Sci. Rep. No. 1, Statistical Forecasting Project, Department of Meteorology, Massachusetts Institute of Technology.

Metz, W., 1989: Low-frequency anomalies of atmospheric flow and the effects of cyclone-scale eddies: a canonical correlation analysis. *J. Atmos. Sci.*, **46,** 1026–1041.

Muller, K. E., 1982: Understanding canonical correlation through the general linear model and principal components. *Amer. Stat.*, **36,** 342–354.

Nicholls, N., 1987: The use of canonical correlation to study teleconnections. *Mon. Wea. Rev.*, **115,** 393–399.

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

Obukhov, A. M., 1960: On statistical orthogonal expansions in empirical functions. *Izvest. Geophys. Ser.* (Eng. Trans), 288–291.

Pearson, K., 1902: On lines and planes of closest fit to system of points in space. *Philos. Mag.*, **6,** 559–572.

Preisendorfer, R. W., 1988: *Principal Component Analysis in Meteorology and Oceanography,* C. Mobley, Ed., Elsevier, 418 pp.

Prohaska, J., 1976: A technique for analyzing the linear relationships between two meteorological fields. *Mon. Wea. Rev.*, **104,** 1345–1353.

Spearman, C., 1904a: The proof and measurement of the association between two things. *Am. J. Psych.*, **15,** 72–101.

——, 1904b: General intelligence, objectively determined and measured. *Am. J. Psych.*, **15,** 201–293.

Stewart, G. W., 1973: *Introduction to Matrix Computations.* Academic Press, 441 pp.

Strang, G., 1988: *Linear Algebra and its Applications.* Harcourt, Brace, and Jovanovitch, 505 pp.

Wallace, J. M., C. Smith and C. S. Bretherton, 1990: Singular value decomposition of sea-surface temperature and 500-mb height anomalies. *J. Climate.*

——, ——, and Q. Jiang, 1990: Spatial patterns of atmosphere–ocean interaction in the Northern winter. *J. Climate,* 3, 990–998.