Singular Value Decomposition Analysis and Canonical Correlation Analysis

STEVE CHERRY

Department of Mathematical Sciences, Montana State University, Bozeman, Montana

(Manuscript received 22 July 1994, in final form 1 February 1996)

ABSTRACT

The goal of singular value decomposition analysis (SVD) and canonical correlation analysis (CCA) is to isolate important coupled modes between two geophysical fields of interest. In this paper the relationship between SVD and CCA is clarified. They should be considered as two distinct methods (possibly) suitable for answering two different questions. Some problems associated with interpreting results of both SVD and CCA are discussed. Both methods have a high potential to produce spurious spatial patterns. Caution is always called for in interpreting results from either method.

1. Introduction

Bretherton et al. (1992, hereafter BSW) discuss methods of "isolating important coupled modes of variability between time series of two fields." A comparison of two methods, Singular-Value Decomposition Analysis (SVD) and Canonical Correlation Analysis (CCA), constitutes a majority of the paper. SVD is based on a singular value decomposition of the matrix whose elements are sample covariances between observations made at different grid points in two geophysical fields. The first use of SVD in climatology was apparently by Prohaska (1976). Other examples of its use can be found in Lanzante (1984), Wallace et al. (1992; hereafter WSB), Hsu (1994), and Lau and Nath (1994).

SVD is known, though not by that name, and practiced in other fields. In the social sciences it is one of a class of methods of matching matrices. Van de Geer (1984) referred to it as the MAXDIFF criterion. In an unpublished manuscript, Muller (1982, personal communication) refers to SVD as Canonical Covariance Analysis, which is perhaps a more descriptive name. Tucker (1958) also discusses SVD as a method of finding common factors in two batteries of tests presented to the same group of subjects (interbattery factor analysis) and comments on the fact that the method results in linear combinations with maximum covariance. SVD is also used in ecology where it is known as co-inertia analysis (Dolédec and Chessel 1994).

SVD is recommended by BSW on the grounds of its simplicity and ease of interpretation of its results. However, in a recent paper, Newman and Sardeshmukh (1995) raised questions about the usefulness of SVD. In particular, they argued that SVD was capable of detecting coupled patterns only under very special circumstances.

This paper has two goals. One is to show that SVD and CCA should not be considered as competing techniques; they are different techniques with different goals. The second goal is to illustrate some potential difficulties with interpreting the results of both methods. Such difficulties are well documented in the statistical literature for CCA (Kendall 1975; Rencher 1992). SVD shares many of those problems and has a high potential to produce spurious patterns and correlations.

To set the stage, suppose there are two geophysical fields. BSW called these the left and right fields. There are $i = 1, \cdots, N_s$ grid points in the left field and $j = 1, \cdots, N_z$ grid points in the right field. Let $\mathbf{s} = (s_1, s_2, \cdots, s_{N_s})'$ and $\mathbf{z} = (z_1, z_2, \cdots, z_{N_z})'$ be $N_s \times 1$ and $N_z \times 1$ *random* vectors. The notation ($'$) means the transpose of a vector or matrix. Each element of $\mathbf{s}$ and $\mathbf{z}$ is a random variable of interest; for example, s_i might represent sea surface temperature at grid point i in the left field and z_j might be the 500-mb height anomaly at the jth grid point in the right field. For convenience it will be assumed that these vectors have mean 0; that is, $\langle s_i \rangle = \langle z_j \rangle = 0$ for all i and j.

Suppose these two random fields are observed over T time units. Then at each grid point there is a time series of T observations. Let $\mathbb{S}$ be a $T \times N_s$ data matrix in which the ith column contains the T observations of s_i and let $\mathbb{Z}$ be a $T \times N_z$ data matrix whose jth column contains the T observations of z_j.

Corresponding author address: Dr. Steve Cherry, Department of Mathematical Sciences, Montana State University, Bozeman, MT 59717-0240.

Let

$$\mathbf{C}_{ss} = \frac{1}{(T-1)}\mathbf{S}'\mathbf{S} \quad (N_s \times N_s)$$

$$\mathbf{C}_{zz} = \frac{1}{(T-1)}\mathbf{Z}'\mathbf{Z} \quad (N_z \times N_z)$$

$$\mathbf{C}_{sz} = \frac{1}{(T-1)}\mathbf{S}'\mathbf{Z} \quad (N_s \times N_z)$$

$$\mathbf{C}_{zs} = \mathbf{C}'_{sz}$$

denote the indicated sample covariance matrices based on recorded observations in the two fields. The combined sample covariance matrix is

$$\mathbf{C}^{(c)} = \begin{pmatrix} \mathbf{C}_{ss} & \mathbf{C}_{sz} \\ \mathbf{C}_{zs} & \mathbf{C}_{zz} \end{pmatrix}.$$

Section 2 discusses CCA and SVD, respectively. Section 3 contains a discussion, and some conclusions are presented in section 4.

2. CCA and SVD

CCA is a multivariate statistical technique whose goal is to examine the strength of the linear association between two sets of random variables or, in the climatological context, between two geophysical fields. CCA, and some modified versions of it, have been used for this purpose in the past in climatology (Glahn 1968; Nicholls 1987; Barnett and Preisendorfer 1987). The subsequent presentation essentially follows that in Seber (1984, 256–268) and Morrison (1990, 305–309).

There is considerable evidence about the association between **s** and **z** in $\mathbf{C}_{sz}$. However, large (small) values in $\mathbf{C}_{sz}$ do not necessarily mean strong (weak) associations. An adequate study of the association between fields requires a consideration of the within-field correlation structure (Morrison 1990). CCA attempts to quantify the association between **s** and **z** by finding a reduced set of variables derived from them with the highest possible correlations. Because correlation is a linear concept, the most natural way to derive these new variables is by taking linear combinations of **s** and **z**, respectively.

To accomplish this reduction the following problem is solved. Find linear combinations of the form

$$x_1 = \mathbf{a}_1'\mathbf{s} \quad y_1 = \mathbf{b}_1'\mathbf{z}$$
$$x_2 = \mathbf{a}_2'\mathbf{s} \quad y_2 = \mathbf{b}_2'\mathbf{z}$$
$$\vdots \qquad \vdots$$
$$x_d = \mathbf{a}_d'\mathbf{s} \quad y_d = \mathbf{b}_d'\mathbf{z}$$

that have the properties that x_1 and y_1 have maximum sample correlation (r_1), x_2 and y_2 have maximum sample correlation (r_2) among all linear combinations that are uncorrelated with x_1 and y_1, and so on for all $d = \min(N_s, N_z)$ pairs. The solution to this problem is such that x_i is uncorrelated with x_j, y_i is uncorrelated with y_j, and x_i is uncorrelated with y_j for $i \neq j$. The variables x_i and y_i are referred to as the ith *canonical variables,* the vectors $\mathbf{a}_i$ and $\mathbf{b}_i$ are the ith *canonical vectors,* and r_i is the ith *canonical correlation coefficient.*

One method of solving this problem, described in terms of the sample covariance matrices, is to take the singular-value decomposition of

$$\mathbf{C}_{ss}^{-(1/2)}\mathbf{C}_{sz}\mathbf{C}_{zz}^{-(1/2)}. \tag{1}$$

In this expression, $\mathbf{C}_{ss}^{-(1/2)}$ is the inverse of

$$\mathbf{C}_{ss}^{(1/2)} = \mathbf{U}\mathbf{E}^{(1/2)}\mathbf{U}',$$

where **U** is a matrix whose columns are the eigenvectors of $\mathbf{C}_{ss}$ and $\mathbf{E}^{(1/2)}$ is a diagonal matrix whose diagonal elements are the positive square roots of the eigenvalues of $\mathbf{C}_{ss}$. Note that $\mathbf{C}_{ss}^{(1/2)}\mathbf{C}_{ss}^{(1/2)} = \mathbf{C}_{ss}$. Term $\mathbf{C}_{zz}^{(1/2)}$ is handled similarly. Writing (1) as

$$\mathbf{C}_{ss}^{-(1/2)}\mathbf{C}_{sz}\mathbf{C}_{zz}^{-(1/2)} = \tilde{\mathbf{A}}\mathbf{D}\tilde{\mathbf{B}}',$$

then the ith columns of $\mathbf{A} = \mathbf{C}_{ss}^{-(1/2)}\tilde{\mathbf{A}}$ and $\mathbf{B} = \mathbf{C}_{zz}^{-(1/2)}\tilde{\mathbf{B}}$ contain the ith pair of canonical vectors and the ith diagonal element of the diagonal matrix **D** contains the square of the ith canonical correlation coefficient.

The solution actually results in T realizations of the d canonical variables: $\mathbf{Sa}_i$ and $\mathbf{Zb}_i$ for $i = 1, \cdots, d$. In the climatological setting these vectors will frequently be a time series and correspond to the CCA expansion coefficients of BSW.

The sample correlation matrix of the canonical variables has the form

$$\mathbf{R}_{\text{CCA}} = \begin{pmatrix} \mathbf{I}_d & \mathbf{R}_{xy} \\ \mathbf{R}_{yx} & \mathbf{I}_d \end{pmatrix},$$

where $\mathbf{I}_d$ is a $d \times d$ identity matrix and $\mathbf{R}_{xy}$ is a $d \times d$ diagonal matrix with $r_1 \geqslant r_2 \geqslant \ldots \geqslant r_d$ on the diagonal. CCA has channeled all the correlation between **s** and **z** through the d canonical variables.

SVD is a method of finding linear combinations of the form $x_i = \mathbf{a}_i'\mathbf{s}$ and $y_i = \mathbf{b}_i'\mathbf{z}$ for $i = 1, \cdots, d$ such that the covariance $c_i = \text{cov}(\mathbf{a}_i'\mathbf{s}, \mathbf{b}_i'\mathbf{z})$ is maximized. Given this data, the problem is to find $\mathbf{a}_i$ and $\mathbf{b}_i$ to maximize $\mathbf{a}_i'\mathbf{C}_{sz}\mathbf{b}_i$ subject to the constraints that $\mathbf{a}_i'\mathbf{a}_i = \mathbf{b}_i'\mathbf{b}_i = 1$ and $\mathbf{a}_i'\mathbf{a}_j = \mathbf{b}_i'\mathbf{b}_j = 0$ for $i \neq j$. A solution is found by taking the singular-value decomposition of $\mathbf{C}_{sz}$ written here as $\mathbf{C}_{sz} = \mathbf{ADB}'$. The variables x_i and y_i are the *singular variables.* The ith columns of **A** and **B**, respectively, contain the ith pair of SVD *vectors* and the ith diagonal element of **D** contains the squared ith *canonical covariance.* As described above for CCA, the solution actually results in T realizations of the d singular variables; $\mathbf{Sa}_i$ and $\mathbf{Zb}_i$, $i = 1, \cdots, d$. These are the SVD expansion coefficients of BSW.

Note that, in general, the singular variable x_i is correlated with $x_1, x_2, \ldots, x_{i-1}$ with similar results holding for y_i. The singular variables x_i are uncorrelated with y_j for $i \neq j$.

The sample covariance matrix of the singular variables is given by

$$\mathbb{C}_{\mathrm{SVD}} = \begin{pmatrix} \mathbb{C}_{xx} & \mathbb{C}_{xy} \\ \mathbb{C}_{yx} & \mathbb{C}_{yy} \end{pmatrix},$$

where $\mathbb{C}_{xy}$ is a $d \times d$ diagonal matrix with the canonical covariances $c_1 \geqslant c_2 \geqslant \cdots \geqslant c_d$ on the diagonal. The $d \times d$ matrices $\mathbb{C}_{xx}$ and $\mathbb{C}_{yy}$ are the sample covariance matrices for the singular variables. The off-diagonal elements of $\mathbb{C}_{xx}$ and $\mathbb{C}_{yy}$ need not be zero. Further, the corresponding correlation matrices $\mathbb{R}_{xx}$ and $\mathbb{R}_{yy}$ need not have zero off-diagonal elements, and the correlation coefficients on the diagonal of $\mathbb{R}_{xy}$ will not be ordered. That is, even though $c_1 \geqslant c_2 \geqslant \cdots \geqslant c_d$ it need not be true, and in general will not be true, that $r_1^* \geqslant r_2^* \geqslant \cdots \geqslant r_d^*$ where

$$r_i^* = c_i/[(\mathbb{a}_i' \mathbb{C}_{ss}\mathbb{a}_i)(\mathbb{b}_i' \mathbb{C}_{zz}\mathbb{b}_i)]^{(1/2)}.$$

Note that the canonical covariances and the SVD vectors computed for standardized data (i.e., correlation matrices) will be different from those computed using nonstandardized data.

The key point is that CCA is a method of analyzing correlation structure, whereas SVD is a method of analyzing covariance structure. Generally, correlation is a better measure of linear association or importance, and so CCA would seem to be more appropriate. However, if the units of measurement and differences in (co)variation are important, then CCA may obscure relevant information, yielding highly correlated but scientifically uninteresting pairs of canonical variables. It is in precisely these circumstances that Muller (1982, personal communication) recommends SVD. He suggests that SVD is more appropriate for analyzing covariance matrices rather than correlation matrices.

BSW discuss some extensions of CCA and SVD that are used to enhance interpretability. The most important of these are the left and right covariance and correlation maps. These maps are constructed from the covariances and correlations between the observations in the respective fields and the expansion coefficients from the same field (homogeneous maps) and the other field (heterogeneous maps). For data with a normalized series of observations, BSW showed that the ith ($i = 1, \cdots, d$) sample left homogeneous and heterogeneous correlation maps for CCA that are equal to $\mathbb{C}_{ss}\mathbb{a}_i$ and $r_i\mathbb{C}_{ss}\mathbb{a}_i$, respectively. For SVD the corresponding maps have the form

$$\mathbb{C}_{ss}\mathbb{a}_i/(\mathbb{a}_i' \mathbb{C}_{ss}\mathbb{a}_i)^{1/2}$$

and

$$c_i\mathbb{a}_k/(\mathbb{b}_i' \mathbb{C}_{zz}\mathbb{b}_i)^{1/2}.$$

The right maps, with obvious notational differences, are computed the same way. The correlations in these vectors are assigned to their corresponding variables (grid points) and maps are produced analogous to maps of EOFs plotted for principal components analysis. The d maps (or some subset of them) are examined for geophysically relevant spatial patterns.

3. Discussion

CCA has a long history of abuse in statistics. The purely mathematical optimization operation always has a solution, but it is often an act of faith or naiveté that results in that solution being given scientific validity. Kendall (1975) writes, ''the difficulties of interpretation are such that not many examples of convincing applications of canonical correlation analysis appear in the literature,'' although one example of a nice application of CCA he does mention is Glahn (1968). There is no reason to suppose that the results of SVD will be any more amenable to interpretation.

Subsequent sections will discuss some of these problems. First a set of examples will be described. The results of those examples will be used to illustrate the potential for SVD to produce spurious results. SVD is sometimes viewed as being superior to CCA because it does not require invertibility of the within field covariance matrices. It seems appropriate to briefly address the problem of singular covariance matrices and small sample sizes; that is also done below. Also, the results of the example presented in BSW will be reexamined.

a. Examples

The examples are based on simulations of seven pairs of fields with different correlation structures. There are 36 stations in both fields located on a one-dimensional transect. In all cases random samples were normally distributed with means of 0 and variances of 1. All of the matrices being considered below are correlation matrices. As stated above, SVD is more properly viewed as a method of analyzing covariance matrices, but the examples are more in line with that presented in BSW, allowing for greater comparability. The conclusions reached below hold for analyses conducted on covariance matrices.

EXAMPLE 1: The observations at the stations are spatially and temporally uncorrelated. A total of 100 independent observations was simulated at each of the 72 stations.

EXAMPLE 2: The observations at the stations are spatially correlated but temporally uncorrelated. There is no signal or coupling between the two fields. The correlation structure is described by what geostatisticians refer to as an exponential covariance function,

$$c(h) = v \exp(-3h/r).$$

In this equation h is the spatial lag, $v = c(0)$ is the variance, and r is defined to be the lag at which the observations become spatially uncorrelated (Issaks and Srivastava 1989). The grid points are assumed to be one unit apart so h ranged in value from 0 to 35. The variance was assumed to be 1, and r was assumed to be 21. Data for the left field was simulated as follows. A 36×36 lag matrix $\mathbf{H}$ with elements $h_{ij} = |i - j|$ was formed. The covariance function was applied to the elements of this matrix creating the spatial covariance matrix $\Sigma_{\mathbf{H}}$. This positive-definite matrix can be decomposed as the product of a unique lower triangular matrix $\mathbf{L}'$ and upper triangular matrix $\mathbf{L}$ such that $\Sigma_{\mathbf{H}} = \mathbf{L}'\mathbf{L}$. This is referred to as a Cholesky Decomposition (Seber 1984, 522). If $\mathbf{u}$ is a 36×1 random vector with elements u_i being normally distributed with means of 0 and variances of 1, then $\mathbf{L}'\mathbf{u}$ is multivariate normally distributed with mean 0 and covariance matrix $\Sigma_{\mathbf{H}}$. This method was used to transform random samples of size 36 drawn from a normally distributed population with mean 0 and variance 1 into Gaussian realizations of the left spatial field with the indicated spatial correlation structure. The process was repeated 100 times giving 100 independent observations at the 36 grid points in each field.

EXAMPLE 3: This example is identical to example 2 except that there were only 20 observations at each station.

EXAMPLE 4: The observations at each station are spatially correlated as described above, but there is also a deterministic signal shared by the two fields. The signal was incorporated in the left field by adding a signal matrix to the data matrix. The signal matrix was constructed as follows. Let ϕ be a 36×1 vector with elements given by

$$\phi_i = \exp[-(y_i^2/2)] \quad i = 1, \cdots, 36$$

and

$$y_i = \frac{(i-1)}{7} - 1.$$

Let $\mathbf{f}$ be a $T \times 1$ vector with $f_j = (2)^{1/2} \sin(t_j)$ and let η be a positive scalar. Then the $T \times 36$ signal matrix is equal to $\eta \mathbf{f} \phi'$. The same process was used to simulate the right field, except the signal matrix was subtracted from the resulting data matrix. For this example, $\eta = 0.4$ and $T = 100$.

EXAMPLE 5: This example is identical to example 4 except $T = 20$.

EXAMPLE 6: This example is identical to example 4 except that $\eta = 1$.

EXAMPLE 7: This example is identical to example 6 except that $T = 20$.

In examples 1, 2, and 3 the true cross-covariance matrix Σ_{sz} is a zero matrix. In the other examples Σ_{sz} is nonzero.

In WSB and Hsu (1994) an important result that is reported is the correlation between the singular variables. In terms of the notation used in this paper, the correlation between the ith pair can be denoted by $r(x_i, y_i)$. Table 1 shows the results of calculating such correlations for the first five modes in the seven examples described above.

First, compare example 2 with 3, 4 with 5, and 6 with 7. In each case when there are only 20 observations per grid point (examples 3, 5, and 7), the correlations are higher than when there are 100 observations (examples 2, 4, and 6). SVD is finding *more* linear structure when there is *less* information. Second, compare examples 2, 4, and 6 with one another and 3, 5, and 7 with one another. There is no appreciable difference between the correlations within each set, yet there is no signal in examples 2 and 3, a relatively weak signal in examples 4 and 5, and a stronger signal in examples 6 and 7. The only exception is in the first mode of example 6 where there are 100 observations and a relatively strong signal. Note that example 1 has the highest correlations of all and that is the example with no spatial correlation and no signal. There is, however, a good deal of spurious linear structure for SVD to exploit.

As a further illustration of the potential seriousness of the problem of small sample sizes, 100 realizations of the random process described in example 1 were simulated, but with 40 observations at each of 150 grid points in each field. These simulated datasets were subjected to an SVD analysis. The means of the correlations between the first five pairs of singular variables were determined, and these were all ~0.95.

These examples and the information in Table 1 are open to criticism. SVD involves more than determining the correlation between pairs of expansion coefficients. But the point made by the results in Table 1 is still a valid one.

b. Correlation maps

Some of the conclusions drawn in WSB and Hsu (1994) are based on a simultaneous comparison of the ith homogeneous and heterogeneous correlation maps.

TABLE 1. The mean correlations between the singular variables from the first five modes in seven examples described in the text. The means are based on 100 simulations of each example.

	Examples						
	1	2	3	4	5	6	7
$r(x_1, y_1)$	0.74	0.32	0.61	0.33	0.60	0.57	0.67
$r(x_2, y_2)$	0.73	0.29	0.57	0.30	0.58	0.31	0.59
$r(x_3, y_3)$	0.70	0.29	0.59	0.29	0.58	0.30	0.59
$r(x_4, y_4)$	0.69	0.30	0.59	0.28	0.59	0.31	0.60
$r(x_5, y_5)$	0.68	0.30	0.60	0.30	0.60	0.31	0.60

But the ith left and right homogeneous maps are mathematically correlated with the ith left and right heterogeneous maps. For CCA, the correlation between the two maps will *always* be identically equal to 1 even if the two fields are completely independent of one another. The two maps will not be perfectly correlated in SVD but they will be correlated. The seven examples described above contain a wide range of within-field and between-field correlation structures. In the 700 simulated random fields generated in these seven examples 75% of the correlation coefficients between the first left homogeneous and heterogeneous correlation maps were in excess of 0.75. With such strong correlations there are going to be recognizable patterns, but the patterns do not necessarily have any geophysical interpretation. In fact, the highest correlations were observed between the two fields in example 1 with the *minimum* of the 100 correlations being equal to 0.93.

The homogeneous maps also are subject to misinterpretation, both in CCA and SVD. The ith left and right homogeneous maps are correlated with the first, second, up to $(i-1)$th maps, leading to the possibility of spurious patterns in a manner analogous to that seen for homogeneous and heterogeneous maps above. The stronger the correlation between the singular variables the more of a problem this will be.

The potential for this is illustrated in Table 2. These results are based on example 2. There is no common signal, although the two fields do have the same spatial correlation structure. Table 2 shows the minimum, maximum, and quartiles of the absolute value of the 100 correlation coefficients estimating the correlation between the first and second left homogeneous correlation maps. The results for both SVD and CCA are shown. Both sets of maps show a tendency to be confounded with one another, even in the absence of any signal or coupling between the two fields.

The implication of the results presented here is that the "coupled" spatial patterns seen in the various correlation maps may be due as much to mathematics as to geophysics, regardless of whether or not there is a geophysical linkage between the two fields.

c. Small sample sizes

The increased probability of spurious correlations with small sample sizes was discussed above. But there is another problem of small sample sizes that needs to be addressed. One of the supposed strengths of SVD relative to CCA is that SVD is possible when the number of observations is less than the number of variables (grid points), that is, when either $\mathbb{C}_{ss}$ or $\mathbb{C}_{zz}$ (or both) are singular. As motivated in section 2, CCA requires invertibility of the within-field covariance matrices. Solutions based on generalized inverses can still be found (Muller 1982), but BSW point out correctly that such solutions are more difficult to interpret.

As pointed out above, SVD and CCA are two different methods whose use does not necessarily overlap.

TABLE 2. The minimum, maximum, and quartiles of the absolute values of 100 correlation coefficients estimating the correlation between the first and second left homogeneous correlation maps. Maps were produced from both CCA and SVD analyses of simulated data in example 2.

	0%	25%	50%	75%	100%
SVD	0.003	0.20	0.47	0.67	0.93
CCA	0.01	0.22	0.48	0.68	0.91

If analysis of covariance structure is the goal, then CCA will not be appropriate even if the number of grid points is greater than the number of variables; and if analysis of the correlation structure is important, then SVD will not be appropriate even if the within-field covariance matrices are singular. In other words, SVD should not be viewed as a method of doing CCA when $\mathbb{C}_{ss}$ and/or $\mathbb{C}_{zz}$ are singular. Doing so would mean that instead of choosing a generalized inverse to find a CCA solution, one is simply assuming that $\mathbb{C}_{ss}$ and $\mathbb{C}_{zz}$ are proportional to identity matrices.

d. BSW's comparative example

BSW presented an example in which SVD performed much better than CCA. Their example is similar to examples 4 and 6. Their goal was to identify a spatial signal ϕ. BSW argued that the SVD vectors and the homogeneous correlation maps in CCA were the most appropriate estimators of the signal. They described a method of comparing these patterns with the true signal and attributed the poor performance of CCA in the simulations involving finite time periods to sampling variation. But CCA performs so poorly that it seems something more than mere sampling variability is to blame.

Recall that CCA can be carried out by performing a singular-value decomposition of $\mathbb{C}_{ss}^{-(1/2)}\mathbb{C}_{sz}\mathbb{C}_{zz}^{-(1/2)}$. In a sense then, CCA can be considered a form of weighted SVD. The weighting takes into account the within-field covariance structure. In general, this is a desirable thing to do. However, if the signal through which two fields are coupled is masked by the within-field covariance structure, then the homogeneous correlation maps in CCA may not be able to isolate it. This is essentially the case in the BSW example. The true signal is confounded by a temporal nuisance signal, and the confounding is worse in the within-field covariance matrices than in the between-field covariance matrices. In such a situation, CCA may be doing a good job of estimating the true canonical vectors, variables, and correlation coefficients, and so its poor performance would not be due to sampling variability. It is possible to unweight the patterns in such a way that CCA can be used to identify the signal, however. This unweighting is accomplished by premultiplying the left and right homogeneous correlation maps by $\mathbb{C}_{ss}^{(1/2)}$ and $\mathbb{C}_{zz}^{(1/2)}$,

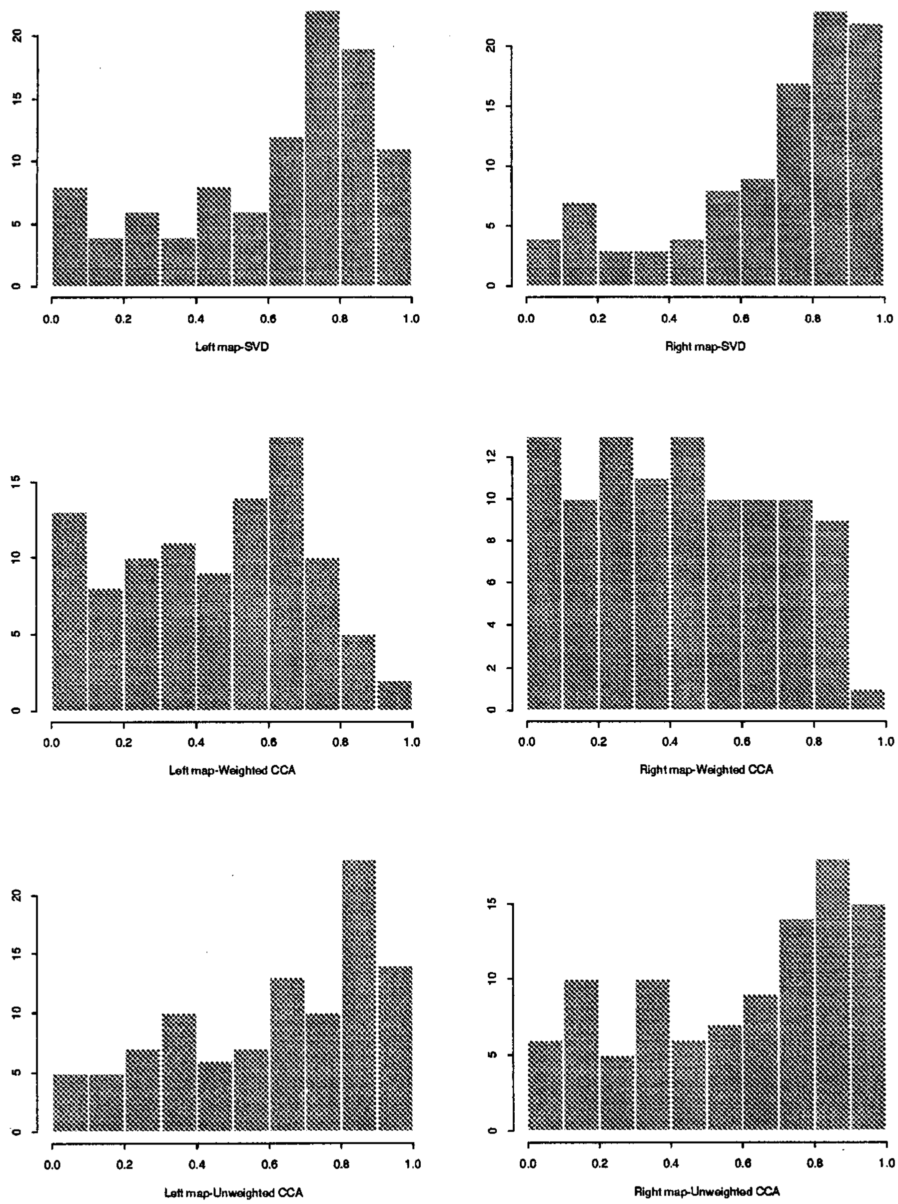

FIG. 1. Histograms of 100 correlation coefficients measuring the correlation between patterns and a deterministic signal. The patterns are from an SVD and CCA analysis of 100 simulated datasets.

respectively. The unweighted patterns then have the form $\mathbf{C}_{ss}^{(3/2)}\mathbf{a}_1$ and $\mathbf{C}_{zz}^{(3/2)}\mathbf{b}_1$.

A total of 100 realizations of example 4 were simulated to evaluate how well these unweighted patterns did in identifying the signal. Left and right SVD vectors, CCA homogeneous correlation maps, and unweighted CCA homogeneous correlation maps were determined and the correlation between these vectors and the true signal was calculated. This is a different method of evaluating performance than that used in BSW; however, it is not an unreasonable one and is easier to implement. Figure 1 shows histograms of the absolute value of these correlation coefficients. The unweighted CCA patterns do a better job of identifying the signal.

4. Summary and conclusions

The goal of SVD and CCA is to isolate important coupled modes between two geophysical fields of interest. SVD is based on an examination of the between-field cross-covariance piece of the combined covariance matrix and may be appropriate when it is the covariances that are of interest. It is suggested that SVD be viewed as a method for analyzing covariance matrices and not correlation matrices.

There is a potential for spurious patterns and correlations to show up in SVD, just as there is for CCA. This potential is increased for small sample sizes. It is worth noting that Kendall (1975) commented on the increased potential for spurious patterns in CCA when the observations are not independent of one another (e.g., if the observations represent a temporally autocorrelated time series). This is likely to be true for SVD.

SVD has been justified on the basis of its simplicity and the supposed ease of interpretation of its results. The results of SVD can be as difficult to interpret as those of any other technique and ease of use is not a justification if the method is not appropriate for the particular data being analyzed. This paper has focused on the potential for SVD (and CCA) to produce highly correlated and seemingly coupled patterns even when there is no relationship. Further, even when there is coupling, the patterns seen in the correlation maps may be due to mathematics rather than geophysics. It complements the results of Newman and Sardeshmukh (1995), who showed that SVD will identify couplings only in special circumstances. This latter limitation of SVD is a consequence of the orthogonality constraints of the maximization procedure. It is clear that used incorrectly or carelessly SVD, for all its simplicity, will produce misleading results.

Acknowledgments. The author would like to thank K. Trenberth, R. Boik, F. Zwiers, D. Shea, R. Madden, R. Katz, J. Hurrell, A. Dempster, I. T. Jolliffe, and J. M. Wallace for advice and comments. Acknowledgment does not imply agreement with the opinions of the author. This work was supported by the National Science Foundation through the Geophysical Statistics Project at the National Center for Atmospheric Research in Boulder, Colorado.

REFERENCES

Barnett, T. P., 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.

Bretherton, C. S., C. Smith, and J. M. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate data. *J. Climate*, **5**, 541–560.

Dolédec, S., and D. Chessel, 1994: Co-inertia analysis: An alternative method for studying species-environment relationships. *Freshwater Biol.*, **31**, 277–294.

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

Hsu, H., 1994: Relationship between tropical heating and global circulation: Interannual variability. *J. Geophys. Res.*, **99**, 10 473–10 489.

Isaaks, E. H., and R. M. Srivastava, 1989: *An Introduction to Applied Geostatistics.* Oxford University Press, 561 pp.

Kendall, M. G., 1975: *Multivariate Analysis.* Griffen, 210 pp.

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.

Lau, N., and M. J. Nath, 1994: A modeling study of the relative roles of tropical and extratropical SST anomalies in the variability of the global atmosphere–ocean system. *J. Climate*, **7**, 1184–1207.

Morrison, D. F., 1990: *Multivariate Statistical Methods.* 3d ed. McGraw-Hill, 495 pp.

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

Newman, M., and P. D. Sardeshmukh, 1995: A caveat concerning singular value decomposition. *J. Climate*, **8**, 352–360.

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

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

Rencher, A. C., 1992: Interpretation of canonical discriminant functions, canonical variates, and principal components. *Amer. Stat.*, **46**, 217–225.

Seber, G. A. F., 1984: *Multivariate Observations.* John Wiley, 686 pp.

Tucker, L. R., 1958: An interbattery method of factor analysis. *Psychometrika*, **23**, 111–136.

van de Geer, J. P., 1984: Linear relations among *k* sets of variables. *Psychometrika*, **49**, 79–94.

Wallace, J. M., C. Smith, and C. S. Bretherton, 1992: Singular-value decomposition of sea surface temperature and 500-mb height anomalies. *J. Climate*, **5**, 561–576.