Chapter 14

Multivariate Analysis of Vector Pairs

14.1. FINDING COUPLED PATTERNS: CCA, MCA, AND RA

This chapter describes three allied multivariate statistical methods: canonical correlation analysis (CCA), maximum covariance analysis (MCA), and redundancy analysis (RA). All three methods find sequences of pairwise relationships between two multivariate data sets. These relationships are constructed in terms of linear combinations, or patterns, in one data set that maximize interrelationships with linear combinations in the other data set. The three methods differ according to the criterion used to define the nature of the interrelationships that are maximized. These approaches thus bear some similarity to PCA (Chapter 13), which searches for patterns within a single multivariate data set that successively represent maximum amounts of its variations.

Denote data vectors from the two multivariate data sets whose relationships are to be characterized as $\boldsymbol{x}$ and $\boldsymbol{y}$. The three methods CCA, MCA, and RA all extract relationships between pairs of data vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ that are summarized in their joint covariance matrix. To compute this matrix, the two centered data vectors are concatenated into a single vector $\boldsymbol{c}'^{\mathrm{T}} = [\boldsymbol{x}'^{\mathrm{T}}, \boldsymbol{y}'^{\mathrm{T}}]$. This partitioned vector contains $I + J$ elements, the first I of which are the elements of $\boldsymbol{x}'$, and the last J of which are the elements of $\boldsymbol{y}'$. The $((I+J) \times (I+J))$ covariance matrix of $\boldsymbol{c}'$, $[S_{\mathrm{C}}]$, is then partitioned into four blocks, in a manner similar to the partitioned covariance matrix in Equation 11.81 or the correlation matrix in Figure 13.5. That is,

$$[S_C] = \frac{1}{n-1}[C']^T[C'] = \begin{bmatrix} [S_{x,x}] & [S_{x,y}] \\ [S_{y,x}] & [S_{y,y}] \end{bmatrix}. \tag{14.1}$$

Each of the n rows of the $(n \times (I+J))$ matrix $[C']$ contains one observation of the vector $\boldsymbol{x}'$ and the corresponding observation of the vector $\boldsymbol{y}'$, with the primes indicating centering of the data by subtraction of each of the respective sample mean vectors. The $(I \times I)$ matrix $[S_{x,x}]$ is the variance–covariance matrix of the I variables in $\boldsymbol{x}$. The $(J \times J)$ matrix $[S_{y,y}]$ is the variance–covariance matrix of the J variables in $\boldsymbol{y}$. The matrices $[S_{x,y}]$ and $[S_{y,x}]$ contain the covariances between all combinations of the elements of $\boldsymbol{x}$ and the elements of $\boldsymbol{y}$, and are related according to $[S_{x,y}] = [S_{y,x}]^{\mathrm{T}}$.

The three methods all find linear combinations of the original variables,

$$v_m = \boldsymbol{a}_m^T \boldsymbol{x}' = \sum_{i=1}^{I} a_{m,i}\, x_i', \quad m = 1, \ldots, \min(I, J); \tag{14.2a}$$

Statistical Methods in the Atmospheric Sciences. https://doi.org/10.1016/B978-0-12-815823-4.00014-6

and

$$w_m = \boldsymbol{b}_m^T \boldsymbol{y}' = \sum_{j=1}^{J} b_{m,j}\, y_j', \quad m = 1, \ldots, \min\,(I, J), \tag{14.2b}$$

by projecting them onto coefficient vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$. These coefficient vectors are defined differently for the three methods, on the basis of the relationships between the derived variables v_m and w_m, as characterized by the information in the joint covariance matrix in Equation 14.1.

CCA, MCA, and RA have been most widely applied to geophysical data in the form of spatial fields. In this setting the vector $\boldsymbol{x}$ often contains observations of one variable at a collection of gridpoints or locations, and the vector $\boldsymbol{y}$ contains observations of a different variable at a set of locations that may (in which case $I = J$) or may not be (in which case usually $I \neq J$) the same as those represented in $\boldsymbol{x}$. Typically the data consist of time series of observations of the two fields. When individual observations of the fields $\boldsymbol{x}$ and $\boldsymbol{y}$ are made simultaneously, these analyses can be useful in diagnosing aspects of the coupled variability of the two fields. When observations of $\boldsymbol{x}$ precede observations of $\boldsymbol{y}$ in time, the methods are natural vehicles for construction of linear statistical forecasts of the $\boldsymbol{y}$ field using the $\boldsymbol{x}$ field as a predictor.

14.2. CANONICAL CORRELATION ANALYSIS (CCA)

14.2.1. Properties of CCA

In CCA, the linear combination, or pattern, vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ in Equation 14.2 are chosen such that each pair of the new variables v_m and w_m, called *canonical variates*, exhibit maximum correlation, while being uncorrelated with the projections of the data onto any of the other identified patterns. That is, CCA identifies new variables that maximize the interrelationships between two data sets in this sense. The vectors of linear combination weights, $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$, are called the *canonical vectors*. The vectors $\boldsymbol{x}'$ and $\boldsymbol{a}_m$ each have I elements, and the vectors $\boldsymbol{y}'$ and $\boldsymbol{b}_m$ each have J elements. The number of pairs, M, of canonical variates that can be extracted from the two data sets is equal to the smaller of the dimensions of $\boldsymbol{x}$ and $\boldsymbol{y}$, that is, $M = \min\,(I, J)$.

The canonical vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ are the unique choices that result in the canonical variates having the properties

$$\mathrm{Corr}(v_1, w_1) \geq \mathrm{Corr}(v_2, w_2) \geq \cdots \geq \mathrm{Corr}(v_M, w_M) \geq 0, \tag{14.3a}$$

$$\mathrm{Corr}(v_k, w_m) = \begin{cases} r_{C_m}, & k = m \\ 0, & k \neq m \end{cases}, \tag{14.3b}$$

$$\mathrm{Corr}(v_k, v_m) = \mathrm{Corr}(w_k, w_m) = 0, \quad k \neq m, \tag{14.3c}$$

and

$$\mathrm{Var}(v_m) = \boldsymbol{a}_m^T [S_{x,x}]\, \boldsymbol{a}_m = \mathrm{Var}(w_m) = \boldsymbol{b}_m^T \left[S_{y,y}\right] \boldsymbol{b}_m = 1, \quad m = 1, \ldots, M. \tag{14.3d}$$

Equation 14.3a states that each of the M successive pairs of canonical variates exhibits no greater correlation than the previous pair. These (Pearson product–moment) correlations between the pairs of canonical variates are called the *canonical correlations*, r_C. The canonical correlations can always be expressed as positive numbers, since either $\boldsymbol{a}_m$ or $\boldsymbol{b}_m$ can be multiplied by -1 if necessary. Equations 14.3b and 14.3c state that each canonical variate is uncorrelated with all the other canonical variates except its specific counterpart in the m^{th} pair. Finally, Equation 14.3d states that each of the canonical variates has unit variance. Some restriction on the lengths of $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ is required for definiteness, and choosing these lengths to yield unit variances for the canonical variates turns out to be convenient for some applications. Accordingly, the joint ($2M \times 2M$) covariance matrix for the resulting canonical variates then takes on the simple and interesting form

$$\mathrm{Var}\left(\left[\frac{\boldsymbol{v}}{\boldsymbol{w}}\right]\right) = \begin{bmatrix} [S_{v,v}] & [S_{v,w}] \\ [S_{w,v}] & [S_{w,w}] \end{bmatrix} = \begin{bmatrix} [I] & [R_C] \\ [R_C] & [I] \end{bmatrix}, \tag{14.4a}$$

where $[R_C]$ is the diagonal matrix of the canonical correlations,

$$[R_C] = \begin{bmatrix} r_{C_1} & 0 & 0 & \cdots & 0 \\ 0 & r_{C_2} & 0 & \cdots & 0 \\ 0 & 0 & r_{C_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & r_{C_M} \end{bmatrix}. \tag{14.4b}$$

The definition of the canonical vectors is reminiscent of PCA, which finds a new orthonormal basis for a single multivariate data set (the eigenvectors of its covariance matrix), subject to a variance maximizing constraint. In CCA, two new bases are defined by the canonical vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$. However, these basis vectors are neither orthogonal nor of unit length. The canonical variates are the projections of the centered data vectors $\boldsymbol{x}'$ and $\boldsymbol{y}'$ onto the canonical vectors and can be expressed in matrix form through the analysis formulae

$$\underset{(M\times 1)}{\boldsymbol{v}} = \underset{(M\times I)}{[A]^{\mathrm{T}}} \underset{(I\times 1)}{\boldsymbol{x}'} \tag{14.5a}$$

and

$$\underset{(M\times 1)}{\boldsymbol{w}} = \underset{(M\times J)}{[B]^{T}} \underset{(J\times 1)}{\boldsymbol{y}'}. \tag{14.5b}$$

Here the columns of the matrices $[A]$ and $[B]$ are the $M = \min(I, J)$ canonical vectors, $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$, respectively. Exposition of how the canonical vectors are calculated from the joint covariance matrix (Equation 14.1) will be deferred to Section 14.2.4.

Unlike the case of PCA, calculating a CCA on the basis of standardized (unit variance) variables yields results that are simple functions of the results from an unstandardized analysis. In particular, because in a standardized analysis the centered variables x_i' and y_j' in Equation 14.2 would be divided by their respective standard deviations, the corresponding elements of the canonical vectors would be larger by factors of those standard deviations. In particular, if $\boldsymbol{a}_m$ is the mth canonical ($I \times 1$) vector for the $\boldsymbol{x}$ variables, its counterpart $\boldsymbol{a}_m{}^*$ in a CCA of the standardized variables would be

$$\boldsymbol{a}_m^* = \boldsymbol{a}_m\,[D_x], \tag{14.6}$$

where the ($I \times I$) diagonal matrix $[D_x]$ (Equation 11.31) contains the standard deviations of the $\boldsymbol{x}$ variables, and a similar equation would hold for the canonical vectors $\boldsymbol{b}_m$ and the ($J \times J$) diagonal matrix $[D_y]$ containing the standard deviations of the y variables. Regardless of whether a CCA is computed using standardized or unstandardized variables, the resulting canonical correlations are the same.

Correlations between the original and canonical variables can be calculated easily. The correlations between corresponding original and canonical variables, sometimes called *homogeneous correlations*, are given by

$$\underset{(1\times I)}{\mathrm{Corr}(v_m, \boldsymbol{x})} = \underset{(1\times I)}{\boldsymbol{a}_m^T} \underset{(I\times I)}{[S_{x,x}]} \underset{(I\times I)}{[D_x]^{-1}} \tag{14.7a}$$

and

$$\underset{(1\times J)}{\mathrm{Corr}(w_m, \boldsymbol{y})} = \underset{(1\times J)}{\boldsymbol{b}_m^T} \underset{(J\times J)}{[S_{y,y}]} \underset{(J\times J)}{[D_y]^{-1}}. \tag{14.7b}$$

These equations specify vectors of correlations, between the mth canonical variable v_m and each of the I original variables x_i; and between the canonical variable w_m and each of the J original variables y_k. Similarly, the vectors of *heterogeneous correlations* between the canonical variables and the "other" original variables are

$$\underset{(1\times J)}{\mathrm{Corr}(v_m, \boldsymbol{y})} = \underset{(1\times I)}{\boldsymbol{a}_m^T} \underset{(I\times J)}{[S_{x,y}]} \underset{(J\times J)}{[D_y]^{-1}} \tag{14.8a}$$

and

$$\underset{(1\times J)}{\mathrm{Corr}(w_m, \boldsymbol{x})} = \underset{(1\times J)}{\boldsymbol{b}_m^T} \underset{(J\times I)}{[S_{y,x}]} \underset{(I\times I)}{[D_x]^{-1}}. \tag{14.8b}$$

The homogeneous correlations indicate correspondence between each canonical variate and the underlying data field, whereas the heterogeneous correlations indicate how well the gridpoints in one field can be specified or predicted by the opposite canonical variate (Bretherton et al., 1992).

The canonical vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ are chosen to maximize correlations between the resulting canonical variates v_m and w_m, but (unlike PCA) may or may not be particularly effective at summarizing the variances of the original variables $\boldsymbol{x}$ and $\boldsymbol{y}$. If canonical pairs with high correlations turn out to represent small fractions of the underlying variability, their physical and practical significance may be limited. Therefore it is often worthwhile to calculate the variance proportions R_m^2 captured by each of the leading canonical variables for its underlying original variable.

How well the canonical variables represent the underlying variability is related to how accurately the underlying variables can be synthesized from the canonical variables. Solving the analysis equations (Equation 14.5) yields the CCA synthesis equations

$$\underset{(I\times 1)}{\boldsymbol{x}'} = \underset{(I\times I)}{\left[\tilde{A}\right]^{-1}} \underset{(I\times 1)}{\boldsymbol{v}} \tag{14.9a}$$

and

$$\underset{(J\times 1)}{\boldsymbol{y}'} = \underset{(J\times J)}{\left[\tilde{B}\right]^{-1}} \underset{(J\times 1)}{\boldsymbol{w}}. \tag{14.9b}$$

If $I = J$ (i.e., if the dimensions of the data vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are equal), then the matrices $[A]$ and $[B]$, whose columns are the corresponding M canonical vectors, are both square. In this case $\left[\tilde{A}\right] = [A]^{\mathrm{T}}$ and $\left[\tilde{B}\right] = [B]^{\mathrm{T}}$ in Equation 14.9, and the indicated matrix inversions can be calculated. If $I \neq J$ then one of the matrices $[A]$ or $[B]$ is not square, and so not invertible. In that case, the last $M - J$ columns of $[A]$ (if $I > J$), or the last $M - I$ columns of $[B]$ (if $I < J$), are filled out with the "phantom" canonical vectors corresponding to the zero eigenvalues, as described in Section 14.2.4.

Equation 14.9 describes the synthesis of individual observations of $\boldsymbol{x}$ and $\boldsymbol{y}$ on the basis of their corresponding canonical variables. In matrix form (i.e., for the full set of n observations), these become

$$\underset{(\mathrm{I\,x\,n})}{[X']^T} = \underset{(\mathrm{I\,x\,I})}{\left[\tilde{A}\right]^{-1}} \underset{(\mathrm{I\,x\,n})}{[V]^T} \tag{14.10a}$$

and

$$\underset{(\mathrm{J\times n})}{[Y']^T} = \underset{(\mathrm{J\times J})}{\left[\tilde{B}\right]^{-1}} \underset{(\mathrm{J\times n})}{[W]^T}. \tag{14.10b}$$

Because the covariance matrices of the canonical variates are $(n–1)^{-1}[V]^T\ [V]=[I]$ and $(n–1)^{-1}$ $[W]^T[W]=[I]$ (Equation 14.4a), substituting Equation 14.10 into Equation 11.30 yields

$$[S_{x,x}] = \frac{1}{n-1}[X']^T[X'] = \left[\tilde{A}\right]^{-1}\left(\left[\tilde{A}\right]^{-1}\right)^T = \sum_{m=1}^{I}\tilde{\boldsymbol{a}}_m\tilde{\boldsymbol{a}}_m^T \tag{14.11a}$$

and

$$[S_{y,y}] = \frac{1}{n-1}[Y']^T[Y'] = \left[\tilde{B}\right]^{-1}\left(\left[\tilde{B}\right]^{-1}\right)^T = \sum_{m=1}^{I}\tilde{\boldsymbol{b}}_m\tilde{\boldsymbol{b}}_m^T, \tag{14.11b}$$

where the canonical vectors with tilde accents indicate columns of the *inverses* of the corresponding matrices. These decompositions are akin to the spectral decompositions (Equation 11.53a) of the two covariance matrices. Accordingly, the proportions of the variances of $\boldsymbol{x}$ and $\boldsymbol{y}$ represented by their mth canonical variables are

$$R_m^2(\boldsymbol{x}) = \frac{\mathrm{tr}\left(\tilde{\boldsymbol{a}}_m\tilde{\boldsymbol{a}}_m^T\right)}{\mathrm{tr}\left([S_{x,x}]\right)} \tag{14.12a}$$

and

$$R_m^2(\boldsymbol{y}) = \frac{\mathrm{tr}\left(\tilde{\boldsymbol{b}}_m\tilde{\boldsymbol{b}}_m^T\right)}{\mathrm{tr}\left([S_{y,y}]\right)}. \tag{14.12b}$$

Example 14.1. CCA of the January 1987 Temperature Data

A simple illustration of the mechanics of a small CCA can be provided by again analyzing the January 1987 temperature data for Ithaca and Canandaigua, New York, given in Table A.1. Let the $I=2$ Ithaca temperature variables be $\boldsymbol{x}=[T_{\max}, T_{\min}]^T$, and similarly let the $J=2$ Canandaigua temperature variables be $\boldsymbol{y}$. The joint covariance matrix $[S_C]$ of these quantities is then the (4×4) matrix

$$[S_C] = \begin{bmatrix} 59.516 & 75.433 & 58.070 & 51.697 \\ 75.433 & 185.467 & 81.633 & 110.800 \\ 58.070 & 81.633 & 61.847 & 56.119 \\ 51.697 & 110.800 & 56.119 & 77.581 \end{bmatrix}. \tag{14.13}$$

This symmetric matrix contains the sample variances of the four variables on the diagonal and the covariances among the variables in the other positions. It is related to the corresponding elements of the

TABLE 14.1 The Canonical Vectors $\boldsymbol{a}_m$ (Corresponding to Ithaca Temperatures) and $\boldsymbol{b}_m$ (Corresponding to Canandaigua Temperatures) for the Partition of the Covariance Matrix in Equation 14.13 With $I = J = 2$

	a_1 (Ithaca)		b_1 (Canandaigua)	a_2 (Ithaca)		b_2 (Canandaigua)
T_{max}	0.0923		0.0946	−0.1618		−0.1952
T_{min}	0.0263		0.0338	0.1022		0.1907
λ_m		0.938			0.593	
$r_{C_m} = \sqrt{\lambda_m}$		0.969			0.770	

Also shown are the eigenvalues λ_m (cf. Example 14.3) and the canonical correlations, which are their square roots.

correlation matrix involving the same variables (see Table 3.5) through the square roots of the diagonal elements: dividing each element by the square root of the diagonal elements in its row and column produces the corresponding correlation matrix. This operation is shown in matrix notation in Equation 11.31.

Since $I = J = 2$, there are $M = 2$ canonical vectors for each of the two data sets being correlated. These are presented in Table 14.1, although the details of their computation will be left until Example 14.3. The first element of each pertains to the respective maximum temperature variable, and the second elements pertain to the minimum temperature variables. The correlation between the first pair of projections of the data onto these vectors, v_1 and w_1, is $r_{C_1} = 0.969$; and the second canonical correlation, between v_2 and w_2, is $r_{C_2} = 0.770$.

Each of the canonical vectors defines a direction in its two-dimensional (T_{max}, T_{min}) data space, but their absolute magnitudes are meaningful only in that they produce unit variances for their corresponding canonical variates. However, the relative magnitudes of the canonical vector elements can be interpreted in terms of which linear combinations of one underlying data vector are most correlated with which linear combination of the other. All the elements of $\boldsymbol{a}_1$ and $\boldsymbol{b}_1$ are positive, reflecting positive correlations among all four temperature variables; although the elements corresponding to the maximum temperatures are larger, reflecting the larger correlation between them than between the minima (Table 3.5). The pairs of elements in $\boldsymbol{a}_2$ and $\boldsymbol{b}_2$ are comparable in magnitude but opposite in sign, suggesting that the next most important pair of linear combinations with respect to correlation relate to the diurnal ranges at the two locations (recall that the signs of the canonical vectors are arbitrary, and chosen to produce positive canonical correlations so that reversing the signs on the second canonical vectors would put positive weights on the maxima and negative weights of comparable magnitudes on the minima).

The time series of the first pair of canonical variables is given by the dot products of $\boldsymbol{a}_1$ and $\boldsymbol{b}_1$ with the pairs of centered temperature values for Ithaca and Canandaigua, respectively, from Table A.1. The value of v_1 for 1 January would be constructed as (33–29.87)(0.0923) + (19–13.00)(0.0263) = 0.447. The time series of v_1 (pertaining to the Ithaca temperatures) would consist of the 31 values (one for each day): 0.447, 0.512, 0.249, −0.449, −0.686, …, −0.041, 0.644. Similarly, the time series for w_1 (pertaining to the Canandaigua temperatures) is 0.474, 0.663, 0.028, −0.304, −0.310, …, −0.283, 0.683. Each of this first pair of canonical variables is a scalar index of the general warmth at its respective location, with more emphasis on the maximum temperatures. Both series have unit sample variance.

The first canonical correlation coefficient, $r_{C_1} = 0.969$, is the correlation between this first pair of canonical variables, v_1 and w_1, and is the largest possible correlation between pairs of linear combinations of these two data sets.

Similarly, the time series of v_2 is 0.107, 0.882, 0.899, −1.290, −0.132, …, −0.225, 0.354 and the time series of w_2 is 1.046, 0.656, 1.446, 0.306, −0.461, …, −1.038, −0.688. Both of these series also have unit sample variance, and their correlation is $r_{C_2} = 0.770$. On each of the $n = 31$ days (the negatives of), these second canonical variates provide an approximate index of the diurnal temperature ranges at the corresponding locations.

The homogeneous correlations (Equation 14.7) for the leading canonical variates, v_1 and w_1, are

$$\mathrm{Corr}\left(v_1, \boldsymbol{x}^T\right) = [0.0923\ 0.0263]\begin{bmatrix} 59.516 & 75.433 \\ 75.433 & 185.467 \end{bmatrix}\begin{bmatrix} 0.1296 & 0 \\ 0 & 0.0734 \end{bmatrix} = [0.969\ \ 0.869] \tag{14.14a}$$

and

$$\mathrm{Corr}\left(w_1, \boldsymbol{y}^T\right) = [0.0946\ \ 0.0338]\begin{bmatrix} 61.847 & 56.119 \\ 56.119 & 77.581 \end{bmatrix}\begin{bmatrix} 0.1272 & 0 \\ 0 & 0.1135 \end{bmatrix} = [0.985\ \ 0.900]. \tag{14.14b}$$

All the four homogeneous correlations are strongly positive, reflecting the strong positive correlations among all four of the variables (see Table 3.5), and the fact that the two leading canonical variables have been constructed with positive weights on all four. The homogeneous correlations for the second canonical variates v_2 and w_2 are calculated in the same way, except that the second canonical vectors $\boldsymbol{a}_2{}^T$ and $\boldsymbol{b}_2{}^T$ are used in Equations 14.14a and 14.14b, respectively, yielding Corr(v_2, $\boldsymbol{x}^{\mathrm{T}}$) = [−0.249, 0.495] and Corr(w_2, $\boldsymbol{y}^{\mathrm{T}}$) = [−0.174, 0.436]. The second canonical variables are less strongly correlated with the underlying temperature variables, because the magnitude of the diurnal temperature range is only weakly correlated with the overall temperatures: wide or narrow diurnal ranges can occur on both relatively warm and cool days. However, the diurnal ranges are evidently more strongly correlated with the minimum temperatures, with cooler minima tending to be associated with large diurnal ranges.

Similarly, the heterogeneous correlations (Equation 14.8) for the leading canonical variates are

$$\mathrm{Corr}\left(v_1, \boldsymbol{y}^T\right) = [.0923\ \ .0263]\begin{bmatrix} 58.070 & 51.697 \\ 81.633 & 110.800 \end{bmatrix}\begin{bmatrix} .1272 & 0 \\ 0 & .1135 \end{bmatrix} = [.955\ \ .872] \tag{14.15a}$$

and

$$\mathrm{Corr}\left(w_1, \boldsymbol{x}^T\right) = [.0946\ \ .0338]\begin{bmatrix} 58.070 & 81.633 \\ 51.697 & 110.800 \end{bmatrix}\begin{bmatrix} .1296 & 0 \\ 0 & .0734 \end{bmatrix} = [.938\ \ .842]. \tag{14.15b}$$

Because of the symmetry of these data (like variables at nearby locations), these correlations are very close to the homogeneous correlations in Equation 14.14. Similarly, the heterogeneous correlations for the second canonical vectors are also close to their homogeneous counterparts: Corr(v_2, $\boldsymbol{y}^{\mathrm{T}}$) = [−0.132, 0.333] and Corr(w_2, $\boldsymbol{x}^{\mathrm{T}}$) = [−0.191, 0.381].

Finally the variance fractions for the temperature data at each of the two locations that are described by the canonical variates depend, through the synthesis equations (Equation 14.9), on the matrices [A] and [B], whose columns are the canonical vectors. Because $I = J$,

$$\left[\widetilde{A}\right] = [A]^T = \begin{bmatrix} 0.0923 & 0.0263 \\ -0.1618 & 0.1022 \end{bmatrix}, \text{ and } \left[\widetilde{B}\right] = [B]^T = \begin{bmatrix} 0.0946 & 0.0338 \\ -0.1952 & 0.1907 \end{bmatrix}; \tag{14.16a}$$

so that

$$\left[\widetilde{A}\right]^{-1} = \begin{bmatrix} 7.466 & -1.921 \\ 11.820 & 6.743 \end{bmatrix}, \text{ and } \left[\widetilde{B}\right]^{-1} = \begin{bmatrix} 7.740 & -1.372 \\ 7.923 & 3.840 \end{bmatrix}. \tag{14.16b}$$

Contributions made by the canonical variates to the respective covariance matrices for the underlying data depend on the outer products of the columns of these inverse matrices (terms in the summations of Equations 14.11), that is,

$$\widetilde{\boldsymbol{a}}_1\widetilde{\boldsymbol{a}}_1^T = \begin{bmatrix} 7.466 \\ 11.820 \end{bmatrix} [7.466 \quad 11.820] = \begin{bmatrix} 55.74 & 88.25 \\ 88.25 & 139.71 \end{bmatrix}, \tag{14.17a}$$

$$\widetilde{\boldsymbol{a}}_2\widetilde{\boldsymbol{a}}_2^T = \begin{bmatrix} -1.921 \\ 6.743 \end{bmatrix} [-1.921 \quad 6.743] = \begin{bmatrix} 3.690 & -12.95 \\ -12.95 & 45.47 \end{bmatrix}, \tag{14.17b}$$

$$\widetilde{\boldsymbol{b}}_1\widetilde{\boldsymbol{b}}_1^T = \begin{bmatrix} 7.740 \\ 7.923 \end{bmatrix} [7.740 \quad 7.923] = \begin{bmatrix} 59.91 & 61.36 \\ 61.36 & 62.77 \end{bmatrix}, \tag{14.17c}$$

$$\widetilde{\boldsymbol{b}}_2\widetilde{\boldsymbol{b}}_2^T = \begin{bmatrix} -1.372 \\ 3.840 \end{bmatrix} [-1.372 \quad 3.840] = \begin{bmatrix} 1.882 & 5.279 \\ 5.279 & 14.75 \end{bmatrix}. \tag{14.17d}$$

Therefore the proportions of the Ithaca temperature variance described by its two canonical variates (Equation 14.12a) are

$$R_1^2(\boldsymbol{x}) = \frac{55.74 + 139.71}{59.52 + 185.47} = 0.798 \tag{14.18a}$$

and

$$R_2^2(\boldsymbol{x}) = \frac{3.690 + 45.47}{59.52 + 185.47} = 0.202, \tag{14.18b}$$

and the corresponding variance fractions for Canandaigua are

$$R_1^2(\boldsymbol{y}) = \frac{59.91 + 62.77}{61.85 + 77.58} = 0.880 \tag{14.19a}$$

and

$$R_2^2(\boldsymbol{y}) = \frac{1.882 + 14.75}{61.85 + 77.58} = 0.120, \tag{14.19b}$$

◇

14.2.2. CCA Applied to Fields

Canonical correlation analysis is usually most interesting for atmospheric data when applied to fields. Here the spatially distributed observations (either at gridpoints or observing locations) are encoded into the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in the same way as for PCA. That is, even though the data may pertain to a two- or three-dimensional field, each location is numbered sequentially and pertains to one element of the corresponding data vector. It is not necessary for the spatial domains encoded into $\boldsymbol{x}$ and $\boldsymbol{y}$ to be the same, and indeed in the applications of CCA that have appeared in the literature they are usually different.

As is the case when using PCA with spatial data, it is often informative to plot maps of the canonical vectors by associating the magnitudes of their elements with the geographic locations to which they pertain. In this context the canonical vectors are sometimes called *canonical patterns*, since the resulting maps show spatial patterns of the ways in which the original variables contribute to the canonical variables. Examining the pairs of maps formed by corresponding vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ can be informative about the nature of the relationship between variations in the data over the domains encoded in $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. Figures 14.2 and 14.3 show examples of maps of canonical vectors.

It can also be informative to plot pairs of maps of the homogeneous (Equation 14.7) or heterogeneous correlations (Equation 14.8). Each of these vectors contains correlations between an underlying data field and one of the canonical variables, and these correlations can also be plotted at the corresponding locations. Figure 14.1 shows one such pair of homogeneous correlation patterns. Figure 14.1a shows the spatial distribution of correlations between a canonical variable v, and the values of the corresponding data $\boldsymbol{x}$ that contains values of average December–February sea-surface temperatures (SSTs) in the north Pacific Ocean. This canonical variable accounts for 18% of the total variance of the SSTs in the data set analyzed (Equation 14.12). Figure 14.1b shows the spatial distribution of the correlations for the corresponding canonical variable w, pertaining to average hemispheric 500 mb heights $\boldsymbol{y}$ during the same winters included in the SST data in $\boldsymbol{x}$. This canonical variable accounts for 23% of the total variance of the winter hemispheric height variations. The correlation pattern in Figure 14.1a corresponds to either cold water in the central north Pacific and warm water along the west coast of North America, or warm water in the central north Pacific and cold water along the west coast of North America. The pattern of 500 mb height correlations in Figure 14.1b is remarkably similar to the PNA pattern (cf. Figures 13.14b and 3.33).

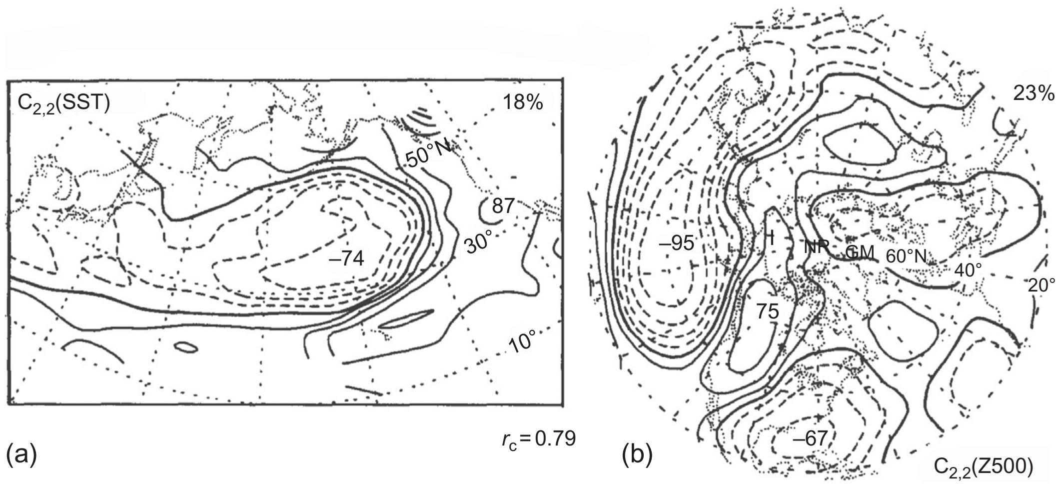

FIGURE 14.1 Homogeneous correlation maps for a pair of canonical variables pertaining to (a) average winter sea-surface temperatures (SSTs) in the northern Pacific Ocean, and (b) hemispheric winter 500 mb heights. The pattern of SST correlation in the left-hand panel (or its negative) is associated with the PNA pattern of 500 mb height correlations shown in the right-hand panel (or *its* negative). The canonical correlation for this pair of canonical variables is 0.79. *From Wallace et al. (1992). © American Meteorological Society. Used with permission.*

The correlation between the two time series v and w is the canonical correlation $r_C = 0.79$. Because v and w are well correlated, these figures indicate that cold SSTs in the central Pacific simultaneously with warm SSTs in the northeast Pacific (relatively large positive v) tend to coincide with a 500 mb ridge over northwestern North America and a 500 mb trough over southeastern North America (relatively large positive w). Similarly, warm water in the central north Pacific and cold water in the northwestern Pacific (relatively large negative v) are associated with the more zonal PNA flow (relatively large negative w).

The sampling properties of CCA may be poor when the available data are few relative to the dimensionality of the data vectors. The result can be that sample estimates for CCA parameters may be unstable (i.e., exhibit large variations from batch to batch) for small samples (e.g., Bretherton et al., 1992; Cherry, 1996). Friederichs and Hense (2003) describe, in the context of atmospheric data, both conventional parametric tests and resampling tests to help assess whether sample canonical correlations may be spurious sampling artifacts. These tests examine the null hypothesis that all the underlying population canonical correlations are zero.

Relatively small sample sizes are common when analyzing time series of atmospheric fields. In CCA, it is not uncommon for there to be fewer observations n than the dimensions I and J of the data vectors, in which case the necessary matrix inversions cannot be computed (see Section 14.2.4). However, even if the sample sizes are large enough to carry through the calculations, sample CCA statistics are erratic unless $n \gg M$. Barnett and Preisendorfer (1987) suggested that a remedy for this problem is to prefilter the two fields of raw data using separate PCAs before subjecting them to a CCA, and this has become a conventional procedure. Rather than directly correlating linear combinations of the fields $\boldsymbol{x}'$ and $\boldsymbol{y}'$, the CCA then operates on the vectors $\boldsymbol{u}_x$ and $\boldsymbol{u}_y$, which consist of the leading principal components of $\boldsymbol{x}$ and $\boldsymbol{y}$. The truncations for these two PCAs (i.e., the dimensions of the vectors $\boldsymbol{u}_x$ and $\boldsymbol{u}_y$) need not be the same, but should be severe enough for the larger of the two to be substantially smaller than the sample size n. Livezey and Smith (1999) provide some guidance for the subjective choices that need to be made in this approach. This combined PCA/CCA approach is not always best and can be inferior if important information is discarded when truncating the PCA. In particular, there is no guarantee that the most strongly correlated linear combinations of $\boldsymbol{x}$ and $\boldsymbol{y}$ will be well related to the leading principal components of one field or the other.

An alternative approach to CCA when $n > M$ is to employ ridge regularization (Section 7.5.1) to the matrices $[S_{x,x}]$ and $[S_{y,y}]$, which allows the CCA computations to proceed because the regularized versions of these matrices are invertible (Cruz-Cano and Lee, 2014; Lim et al., 2012; Vinod, 1976). In addition (or instead, if $n < M$), L1 regularization (Section 7.5.2) can be applied in the computation of the canonical vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ (Hastie et al., 2015; Witten et al., 2009).

14.2.3. Forecasting With CCA

When one of the fields, say $\boldsymbol{x}$, is observed prior to $\boldsymbol{y}$, and some of the canonical correlations between the two are large, it is natural to use CCA either as a purely statistical forecasting method (e.g., Barnston, 1994; Landman and Mason, 2001; Wilks, 2014a) or as a MOS implementation (e.g., Alfaro et al., 2018, Lim et al., 2012). In either case the entire $(I \times 1)$ field $\boldsymbol{x}(t)$ is used to forecast the $(J \times 1)$ field $\boldsymbol{y}(t + \tau)$, where τ is the time lag between the two fields in the training data, which becomes the forecast lead time. In applications with atmospheric data it is typical that there are too few observations n relative to the dimensions I and J of the fields for stable sample estimates (which are especially important for out-of-sample forecasting) to be calculated, even if the calculations can be performed because

$n > \max(I, J)$. It is therefore usual for both the $\boldsymbol{x}$ and $\boldsymbol{y}$ fields to be represented by separate series of truncated principal components, as described in the previous section. However, in order not to clutter the notation in this section, the mathematical development will be expressed in terms of the original variables $\boldsymbol{x}$ and $\boldsymbol{y}$, rather than their principal components $\boldsymbol{u}_x$ and $\boldsymbol{u}_y$.

The basic idea behind forecasting with CCA is straightforward: simple linear regressions are constructed that relate the predictand canonical variates w_m to the predictor canonical variates v_m,

$$w_m = \hat{\beta}_{0,m} + \hat{\beta}_{1,m} v_m, \quad m = 1, \ldots, M. \tag{14.20}$$

Here the estimated regression coefficients are indicated by the β's in order to distinguish clearly from the canonical vectors $\boldsymbol{b}$, and the number of canonical pairs considered can be any number up to the smaller of the numbers of principal components retained for the $\boldsymbol{x}$ and $\boldsymbol{y}$ fields. These regressions are all simple linear regressions that can be computed individually, because canonical variables from different canonical pairs are uncorrelated (Equation 14.3b).

Parameter estimation for the regressions in Equation 14.20 is straightforward also. Using Equation 7.7a for the regression slopes,

$$\hat{\beta}_{1,m} = \frac{n\,\mathrm{Cov}(v_m, w_m)}{n\,\mathrm{Var}(v_m)} = \frac{n\, s_v s_w r_{v,w}}{n\, s_v^2} = r_{v,w} = r_{C_m}, \quad m = 1, \ldots, M. \tag{14.21}$$

That is, because the canonical variates are scaled to have unit variance (Equation 14.3c), the regression slopes are simply equal to the corresponding canonical correlations. Similarly, Equation 7.7b yields the regression intercepts

$$\hat{\beta}_{0,m} = \overline{w}_m - \hat{\beta}_{1,m}\overline{v}_m = \boldsymbol{b}_m^T E(\boldsymbol{y}') + \hat{\beta}_{1,m}\boldsymbol{a}_m^T E(\boldsymbol{x}') = 0, \quad m = 1, \ldots, M. \tag{14.22}$$

That is, because the CCA is calculated from the centered data $\boldsymbol{x}'$ and $\boldsymbol{y}'$ whose mean vectors are both $\mathbf{0}$, the averages of the canonical variables v_m and w_m are also both zero, so that all the intercepts in Equation 14.20 are zero. Equation 14.22 also holds when the CCA has been calculated from a principal component truncation of the original (centered) variables, because $E(\boldsymbol{u}_x) = E(\boldsymbol{u}_y) = \mathbf{0}$.

Once the CCA has been fit, the basic forecast procedure is as follows. First, centered values for the predictor field $\boldsymbol{x}'$ (or its first few principal components, $\boldsymbol{u}_x$) are used in Equation 14.5a to calculate the M canonical variates v_m to be used as regression predictors. Combining Equations 14.20 through 14.22, the $(M \times 1)$ vector of predictand canonical variates is forecast to be

$$\hat{\boldsymbol{w}} = [R_C]\,\boldsymbol{v} = [R_C][A]^T\boldsymbol{x}', \tag{14.23}$$

where $[R_C]$ is the diagonal $(M \times M)$ matrix of the canonical correlations (Equation 14.4b) and the $(I \times M)$ matrix $[A]$ contains the predictor canonical vectors $\boldsymbol{a}_m$. In general, the forecast map $\hat{\boldsymbol{y}}$ will need to be synthesized from its predicted canonical variates using Equation 14.9b, in order to see the forecast in a physically meaningful way. However, in order to be invertible, the matrix $[B]$, whose columns are the predictand canonical vectors $\boldsymbol{b}_m$, must be square. This condition implies that the number of regressions M in Equation 14.20 needs to be equal to the dimensionality of $\boldsymbol{y}$ (or, more usually, to the number of predictand principal components that have been retained), although the dimension of $\boldsymbol{x}$ (or the number of predictor principal components retained) is not constrained in this way. If the CCA has been calculated using predictand principal components $\boldsymbol{u}_y$, the centered predicted values $\hat{\boldsymbol{y}}'$are next recovered with the

PCA synthesis, Equation 13.6. Finally, the full predicted field is produced by adding back its mean vector. If the CCA has been computed using standardized variables, so that Equation 14.1 is a correlation matrix, the dimensional values of the predicted variables need to be reconstructed by multiplying by the appropriate standard deviation before adding the appropriate mean (i.e., by reversing Equation 3.27 or Equation 4.27 to yield Equation 4.29).

Example 14.2. An Operational CCA Forecast System

Canonical correlation is used as one of the elements contributing to the seasonal forecasts produced operationally at the U.S. Climate Prediction Center (Barnston et al., 1999). The predictands are seasonal (three-month) average temperature and total precipitation over the United States, made at lead times of 0.5 through 12.5 months.

The CCA forecasts contributing to this system are modified from the procedure described in Barnston (1994), whose temperature forecast procedure will be outlined in this example. The (59×1) predictand vector $\boldsymbol{y}$ represents temperature forecasts jointly at 59 locations in the conterminous United States. The predictors $\boldsymbol{x}$ consist of global sea-surface temperatures (SSTs) discretized to a 235-point grid, northern hemisphere 700 mb heights discretized to a 358-point grid, and previously observed temperatures at the 59 prediction locations. The predictors are three-month averages also, but in each of the four nonoverlapping three-month seasons for which data would be available preceding the season to be predicted. For example, the predictors for the January–February–March (JFM) forecast, to be made in mid-December, are seasonal averages of SSTs, 700 mb heights, and U.S. surface temperatures during the preceding September–October–November (SON), June–July–August (JJA), March–April–May (MAM), and December–January–February (DJF) seasons, so that the predictor vector $\boldsymbol{x}$ has $4\,(235+358+59) = 2608$ elements. In principle, using sequences of four consecutive predictor seasons allows the forecast procedure to incorporate information about the time evolution of the predictor fields.

Since only $n = 37$ years of training data were available when this system was developed, drastic reductions in the dimensionality of both the predictors and predictands were necessary. Separate PCAs were calculated for the predictor and predictand vectors, which retained the leading six predictor principal components $\boldsymbol{u}_x$ and (depending on the forecast season) either five or six predictand principal components $\boldsymbol{u}_y$. The CCAs for these pairs of principal component vectors yield either $M = 5$ or $M = 6$ canonical pairs. Figure 14.2 shows that portion of the first predictor canonical vector $\boldsymbol{a}_1$ pertaining to the SSTs in the three seasons MAM, JJA, and SON, relating to the forecast for the following JFM. That is, each of these three maps expresses the SST contributions to the six elements of $\boldsymbol{a}_1$ in terms of the original 235 spatial locations, through the corresponding elements of the eigenvector matrix $[E]$ for the predictor PCA. The most prominent feature in Figure 14.2 is the progressive evolution of increasingly negative values in the eastern tropical Pacific, which clearly represents an intensifying El Niño (warm) event when $v_1 < 0$, and development of a La Niña (cold) event when $v_1 > 0$, in the spring, summer, and fall before the JFM season to be forecast.

Figure 14.3 shows the first canonical predictand vector for the JFM forecast, $\boldsymbol{b}_1$, again projected back to physical space at the 59 forecast locations. Because the CCA is constructed to have positive canonical correlations, a developing El Niño yielding $v_1 < 0$ results in a forecast $\hat{w}_1 < 0$ (Equation 14.23). The result is that the first canonical pair contributes a tendency toward relative warmth in the northern United States and relative coolness in the southern United States during El Niño winters. Conversely, this canonical pair forecasts cold in the north and warm in the south for La Niña winters. Evolving SSTs not resembling the patterns in Figure 14.2 would yield $v_1 \approx 0$, resulting in little contribution from the pattern in Figure 14.3 to the forecast. ◇

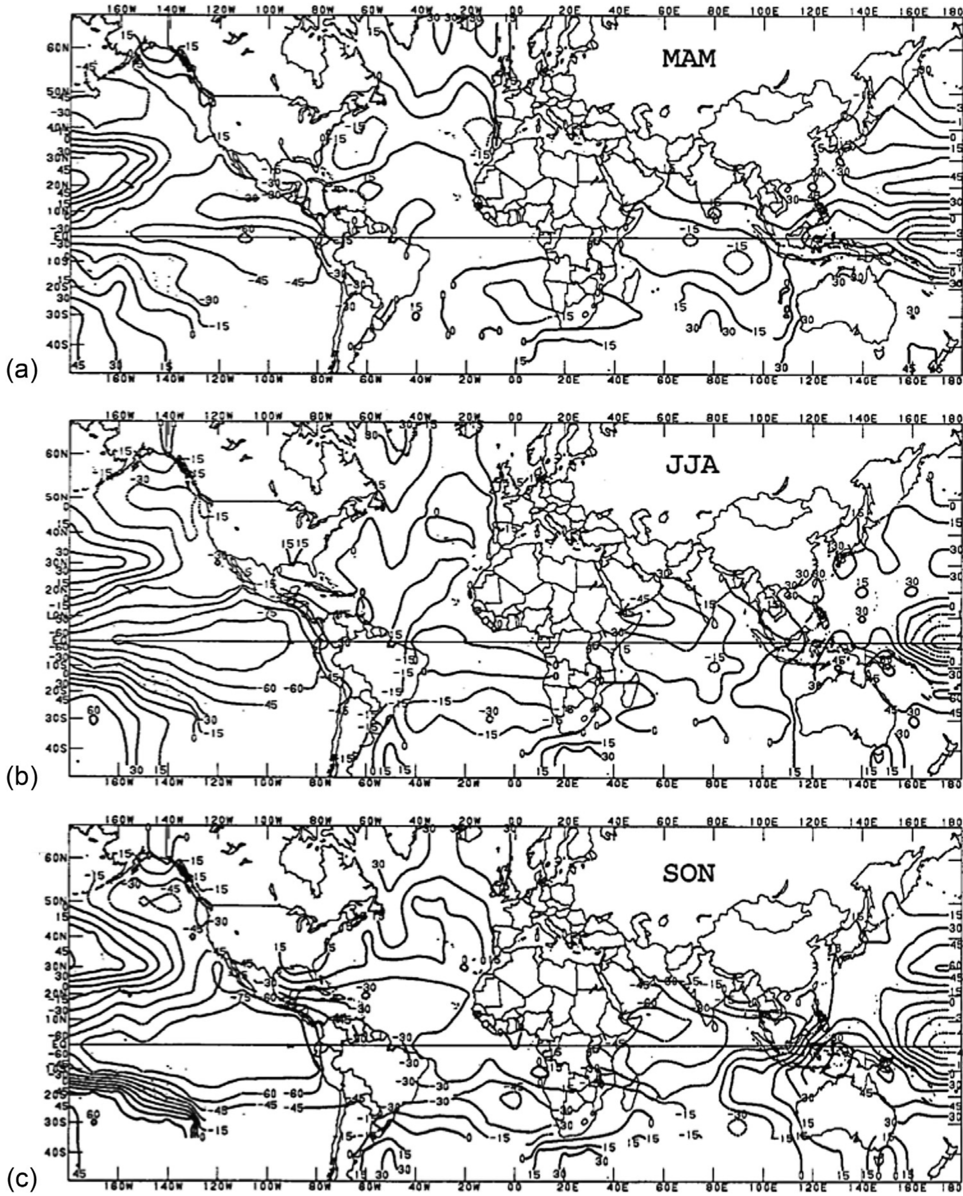

FIGURE 14.2 Spatial displays of portions of the first canonical vector for predictor sea-surface temperatures, in the three seasons preceding the JFM for which U.S. surface temperatures are forecast. The corresponding canonical vector for this predictand is shown in Figure 14.3. *From Barnston (1994). © American Meteorological Society. Used with permission.*

Extending CCA to probabilistic forecasts is relatively straightforward because the canonical variates are uncorrelated with all others except their counterparts in the *m*th pair. In particular, because each of the regressions in Equation 14.20 is independent of the others, the joint covariance matrix for the predictand canonical variates is

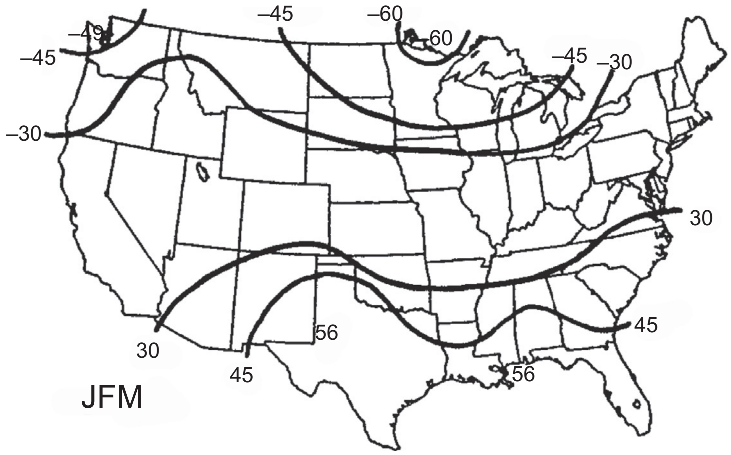

FIGURE 14.3 Spatial display of the first canonical vector for predicted U.S. JFM surface temperatures. A portion of the corresponding canonical vector for the predictors is shown in Figure 14.2. *From Barnston (1994). © American Meteorological Society. Used with permission.*

$$[S_{\hat{w}}] = \begin{bmatrix} s^2_{\hat{w}_1} & 0 & \cdots & 0 \\ 0 & s^2_{\hat{w}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s^2_{\hat{w}_M} \end{bmatrix}, \tag{14.24}$$

where, using Equation 7.23 for the regression prediction variance,

$$s^2_{\hat{w}_m} = s^2_{e_m}\left[1 + \frac{1}{n} + \frac{(v_{0,m} - \overline{v}_m)^2}{\sum_{i=1}^{n}(v_{i,m} - \overline{v}_m)^2}\right] = \left(1 - r^2_{C_m}\right)\left[1 + \frac{1}{n} + \frac{{v_{0,m}}^2}{\sum_{i=1}^{n}{v_{i,m}}^2}\right]. \tag{14.25}$$

Here $v_{0,m}$ denotes the predictor linear combination in implementation, which is not one of the n training data values. The second equality follows because the fraction of variance represented by each regression residual is $1 - r^2_{C_m}$, and each predictand canonical variate has unit variance by construction. The centering of the predictor data implies that the average of the predictor canonical variates is zero.

If it is reasonable to invoke multivariate normality for the forecast errors, predictive distributions are fully defined by the vector mean in Equation 14.9b, and the covariance matrix

$$[S_{\hat{y}'}] = \left([B]^{-1}\right)^T [S_w]\,[B]^{-1}. \tag{14.26}$$

Figure 14.4 shows the 90% forecast probability ellipses derived from bivariate normal predictive distributions for the January–March 1995 SSTs, jointly in the Niño 3 and Niño 4 regions of the equatorial Pacific Ocean, which were computed on the basis of the $I = 4$ leading principal components of Indo-Pacific SSTs observed from 1 to 9 months prior. The forecast uncertainty (areas enclosed by the ellipses) decreases as the lead times become shorter, and the vector means of the distributions (numerals) move toward the observed January–March values (large dot).

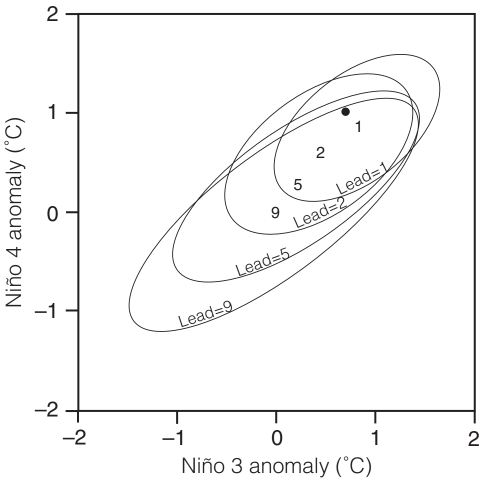

FIGURE 14.4 Ninety percent forecast probability ellipses for joint Niño 3 and Niño 4 forecasts for JFM 1995, at the 1-, 2-, 5-, and 9-month lead times. Numerals show respective forecast means, and the solid dot locates the observed 1995 value. *Modified from Wilks (2014a).*

In problems where a K-dimensional predictand vector has been represented by its leading J principal components, the corresponding predictive mean vector and covariance matrix are

$$\underset{(K\times 1)}{\hat{\mathbf{y}}'} = \underset{(K\times J)}{[E]} \underset{(J\times J)}{\left([B]^T\right)^{-1}} \underset{(J\times 1)}{\mathbf{w}} \tag{14.27}$$

and

$$\underset{(K\times K)}{\left[S_{\hat{\mathbf{y}}'}\right]} = \underset{(K\times J)}{[E]} \underset{(J\times J)}{\left([B]^{-1}\right)^T} \underset{(J\times J)}{[S_{\mathbf{w}}]} \underset{(J\times J)}{[B]^{-1}} \underset{(J\times K)}{[E]^T}, \tag{14.28}$$

where it has been assumed that $I = J$, so that $\left[\tilde{B}\right] = [B]^T$.

Long-lead forecasts of monthly or seasonally averaged quantities are often made with linear multivariate statistical methods such as CCA. Although the dynamics of the climate system are nonlinear, in practice it has been found that nonlinear statistical methods perform no better than the traditional linear statistics for seasonal prediction (Tang et al., 2000; Van den Dool, 2007). Linear multivariate statistical methods also often perform comparably to fully nonlinear dynamical models in seasonal forecasting applications (e.g., Harrison, 2005, Quan et al., 2006, Toth et al., 2007, van den Dool 2007), at substantially reduced computational cost. Possibly this phenomenon occurs because the time averaging inherent in long-lead forecasting renders both the predictors and predictands Gaussian or quasi-Gaussian because of the Central Limit Theorem, which in turn induces linear or quasi-linear relationships among them (Hlinka et al., 2014; Hsieh, 2009; Yuval and Hsieh, 2002).

14.2.4. Computational Considerations

Finding canonical vectors and canonical correlations requires calculating pairs of eigenvectors $\mathbf{e}_m$, corresponding to the $\mathbf{x}$ variables, and eigenvectors $\mathbf{f}_m$, corresponding to the $\mathbf{y}$ variables; together with their corresponding eigenvalues λ_m, which are the same for each pair $\boldsymbol{e}_m$ and $\boldsymbol{f}_m$. There are several computational approaches available to find these $\boldsymbol{e}_m$, $\boldsymbol{f}_m$, and λ_m, $m = 1,\ldots, M$.

Calculating CCA Through Direct Eigendecomposition

One approach is to find the eigenvectors $\boldsymbol{e}_m$ and $\boldsymbol{f}_m$ of the matrices

$$[S_{x,x}]^{-1}[S_{x,y}]\,[S_{y,y}]^{-1}[S_{y,x}] \tag{14.29a}$$

and

$$[S_{y,y}]^{-1}[S_{y,x}]\,[S_{x,x}]^{-1}[S_{x,y}], \tag{14.29b}$$

respectively. The factors in these equations correspond to the definitions in Equation 14.1. Equation 14.29a is dimensioned ($I \times I$), and Equation 14.29b is dimensioned ($J \times J$). The first $M = \min(I, J)$ eigenvalues of these two matrices are equal, and if $I \neq J$, the remaining eigenvalues of the larger matrix are zero. The corresponding "phantom" eigenvectors would fill the extra rows of one of the matrices in Equation 14.9. Equation 14.29 can be difficult computationally because in general these matrices are not symmetric, and algorithms to find eigenvalues and eigenvectors for general matrices are less stable numerically than routines designed specifically for real and symmetric matrices.

The eigenvalue–eigenvector computations are easier and more stable, and the same results are achieved, if the eigenvectors $\boldsymbol{e}_m$ and $\boldsymbol{f}_m$ are calculated from the symmetric matrices

$$[S_{x,x}]^{-1/2}[S_{x,y}]\,[S_{y,y}]^{-1}[S_{y,x}]\,[S_{x,x}]^{-1/2} \tag{14.30a}$$

and

$$[S_{y,y}]^{-1/2}[S_{y,x}]\,[S_{x,x}]^{-1}[S_{x,y}]\,[S_{y,y}]^{-1/2}, \tag{14.30b}$$

respectively. Equation 14.30a is dimensioned ($I \times I$), and Equation 14.30b is dimensioned ($J \times J$). Here the reciprocal square-root matrices must be symmetric (Equation 11.68), and not derived from Cholesky decompositions of the corresponding inverses or obtained by other means. The eigenvalue–eigenvector pairs for the symmetric matrices in Equation 14.30 can be computed using an algorithm specialized to the task, or through the singular value decomposition (Equation 11.72) operating on these matrices. In the latter case, the results are $[E]\,[\Lambda]\,[E]^{\mathrm{T}}$ and $[F]\,[\Lambda]\,[F]^{\mathrm{T}}$, respectively (compare Equations 11.72 and 11.52a), where the columns of $[E]$ are the $\boldsymbol{e}_m$ and the columns of $[F]$ are the $\boldsymbol{f}_m$.

Regardless of how the eigenvectors $\boldsymbol{e}_m$ and $\boldsymbol{f}_m$, and their common eigenvalues λ_m, are arrived at, the canonical correlations and canonical vectors are calculated from them. The canonical correlations are simply the positive square roots of the M nonzero eigenvalues,

$$r_{C_m} = \sqrt{\lambda_m}, \quad m = 1,\ldots,M \tag{14.31}$$

The pairs of canonical vectors are calculated from the corresponding pairs of eigenvectors, using

$$\left.\begin{aligned} \boldsymbol{a}_m &= [S_{x,x}]^{-1/2}\boldsymbol{e}_m \\ \boldsymbol{b}_\mathrm{m} &= [S_{y,y}]^{-1/2}\boldsymbol{f}_m \end{aligned}\right\}, \quad m = 1,\ldots,M. \tag{14.32}$$

Since $||\boldsymbol{e}_m|| = ||\boldsymbol{f}_m|| = 1$, this transformation ensures unit variances for the canonical variates, that is,

$$\mathrm{Var}(v_m) = \boldsymbol{a}_m^T [S_{x,x}]\, \boldsymbol{a}_m = \boldsymbol{e}_m^T [S_{x,x}]^{-1/2} [S_{x,x}]\, [S_{x,x}]^{-1/2} \boldsymbol{e}_m = \boldsymbol{e}_m^T \boldsymbol{e}_m = 1, \tag{14.33}$$

because $[S_{x,x}]^{-1/2}$ is symmetric and the eigenvectors $\boldsymbol{e}_m$ are mutually orthogonal. An obvious analogous equation can be written for the variances $\mathrm{Var}(w_m)$.

Extraction of eigenvalue–eigenvector pairs from large matrices can require large amounts of computing. However, the eigenvector pairs $\boldsymbol{e}_m$ and $\boldsymbol{f}_m$ are related in a way that makes it unnecessary to compute the eigendecompositions of both Equations 14.30a and 14.30b (or, both Equations 14.29a and 14.29b). For example, each $\boldsymbol{f}_m$ can be computed from the corresponding $\boldsymbol{e}_m$ using

$$\boldsymbol{f}_m = \frac{[S_{y,y}]^{-1/2} [S_{y,x}]\, [S_{x,x}]^{-1/2} \boldsymbol{e}_m}{\left\| [S_{y,y}]^{-1/2} [S_{y,x}]\, [S_{x,x}]^{-1/2} \boldsymbol{e}_m \right\|}, \quad m = 1, \ldots, M. \tag{14.34}$$

Here the Euclidean norm in the denominator ensures $||\boldsymbol{f}_m|| = 1$. The eigenvectors $\boldsymbol{e}_m$ can be calculated from the corresponding $\boldsymbol{f}_m$ by reversing their roles, and reversing the matrix subscripts, in this equation.

CCA Via SVD

The special properties of the singular value decomposition (Equation 11.72) can be used to find both sets of the $\boldsymbol{e}_m$ and $\boldsymbol{f}_m$ pairs, together with the corresponding canonical correlations. This is achieved by computing the SVD

$$\underset{(I\times I)}{[S_{x,x}]^{-1/2}} \; \underset{(I\times J)}{[S_{x,y}]} \; \underset{(J\times J)}{[S_{y,y}]^{-1/2}} = \underset{(I\times J)}{[E]} \; \underset{(J\times J)}{[R_C]} \; \underset{(J\times J)}{[F]^T}. \tag{14.35}$$

The left-hand side of Equation 14.35 expresses the covariance between the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ after each has been subjected to the Mahalanobis transformation (Equation 12.21), which has also been called a *whitening transformation* (DelSole and Tippett, 2007; Swenson, 2015). This transformation both decorrelates the variables and scales them to unit variance. Equation 14.35 for the resulting covariance matrix is an instance of Equation 11.89 for the covariances among pairs of linear combinations because the square-root matrices are symmetric. As before the columns of $[E]$ are the $\boldsymbol{e}_m$, the columns of $[F]$ are the $\boldsymbol{f}_m$, and the diagonal matrix $[R_C]$ contains the canonical correlations. Here it has been assumed that $I \geq J$, but if $I < J$ the roles of $\boldsymbol{x}$ and $\boldsymbol{y}$ can be reversed in Equation 14.35. The canonical vectors are calculated as before, using Equation 14.32.

Example 14.3. The Computations behind Example 14.1

In Example 14.1 the canonical correlations and canonical vectors were given, with their computations deferred. Since $I = J$ in this example, the matrices required for these calculations are obtained by quartering $[S_C]$ (Equation 14.13) to yield

$$[S_{x,x}] = \begin{bmatrix} 59.516 & 75.433 \\ 75.433 & 185.467 \end{bmatrix}, \tag{14.36a}$$

$$[S_{y,y}] = \begin{bmatrix} 61.847 & 56.119 \\ 56.119 & 77.581 \end{bmatrix}, \tag{14.36b}$$

and

$$[S_{y,x}] = [S_{x,y}]^T = \begin{bmatrix} 58.070 & 81.633 \\ 51.697 & 110.800 \end{bmatrix}. \tag{14.36c}$$

The eigenvectors $\boldsymbol{e}_m$ and $\boldsymbol{f}_m$, respectively, can be computed either from the pair of asymmetric matrices (Equation 14.29)

$$[S_{x,x}]^{-1}[S_{x,y}]\,[S_{y,y}]^{-1}[S_{y,x}] = \begin{bmatrix} 0.830 & 0.377 \\ 0.068 & 0.700 \end{bmatrix} \tag{14.37a}$$

and

$$[S_{y,y}]^{-1}[S_{y,x}]\,[S_{x,x}]^{-1}[S_{x,y}] = \begin{bmatrix} 0.845 & 0.259 \\ 0.091 & 0.686 \end{bmatrix}; \tag{14.37b}$$

or the symmetric matrices (Equation 14.30).

$$[S_{x,x}]^{-1/2}[S_{x,y}]\,[S_{y,y}]^{-1}[S_{y,x}]\,[S_{x,x}]^{-1/2} = \begin{bmatrix} 0.768 & 0.172 \\ 0.172 & 0.757 \end{bmatrix} \tag{14.38a}$$

and

$$[S_{y,y}]^{-1/2}[S_{y,x}]\,[S_{x,x}]^{-1}[S_{x,y}]\,[S_{y,y}]^{-1/2} = \begin{bmatrix} 0.800 & 0.168 \\ 0.168 & 0.726 \end{bmatrix}. \tag{14.38b}$$

The numerical stability of the computations is better if Equations 14.38a and 14.38b are used, but in either case the eigenvectors of Equations 14.37a and 14.38a are

$$\boldsymbol{e}_1 = \begin{bmatrix} 0.719 \\ 0.695 \end{bmatrix} \quad \text{and} \quad \boldsymbol{e}_2 = \begin{bmatrix} -0.695 \\ 0.719 \end{bmatrix}, \tag{14.39}$$

with corresponding eigenvalues $\lambda_1 = 0.938$ and $\lambda_2 = 0.593$. The eigenvectors of Equations 14.37b and 14.38b are

$$\boldsymbol{f}_1 = \begin{bmatrix} 0.780 \\ 0.626 \end{bmatrix} \quad \text{and} \quad \boldsymbol{f}_2 = \begin{bmatrix} -0.626 \\ 0.780 \end{bmatrix}, \tag{14.40}$$

again with eigenvalues $\lambda_1 = 0.938$ and $\lambda_2 = 0.593$. However, once the eigenvectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ have been computed it is not necessary to compute the eigendecomposition for either Equation 14.37b or Equation 14.38b, because their eigenvectors can also be obtained through Equation 14.34:

$$\boldsymbol{f}_1 = \begin{bmatrix} 0.8781 & 0.1788 \\ 0.0185 & 0.8531 \end{bmatrix}\begin{bmatrix} 0.719 \\ 0.695 \end{bmatrix} \Big/ \left\| \begin{bmatrix} 0.8781 & 0.1788 \\ 0.0185 & 0.8531 \end{bmatrix}\begin{bmatrix} 0.719 \\ 0.695 \end{bmatrix} \right\| = \begin{bmatrix} 0.780 \\ 0.626 \end{bmatrix} \tag{14.41a}$$

and

$$\boldsymbol{f}_2 = \begin{bmatrix} 0.8781 & 0.1788 \\ 0.0185 & 0.8531 \end{bmatrix}\begin{bmatrix} -0.695 \\ 0.719 \end{bmatrix} \Big/ \left\| \begin{bmatrix} 0.8781 & 0.1788 \\ 0.0185 & 0.8531 \end{bmatrix}\begin{bmatrix} -0.695 \\ 0.719 \end{bmatrix} \right\| = \begin{bmatrix} -0.626 \\ 0.780 \end{bmatrix}, \tag{14.41b}$$

since

$$\left[S_{y,y}\right]^{-1/2}\left[S_{y,x}\right]\left[S_{x,x}\right]^{-1/2} = \begin{bmatrix} 0.1960 & -0.0930 \\ -0.0930 & 0.1699 \end{bmatrix} \begin{bmatrix} 58.070 & 81.633 \\ 51.697 & 110.800 \end{bmatrix} \begin{bmatrix} 0.1788 & -0.0522 \\ -0.0522 & 0.0917 \end{bmatrix} = \begin{bmatrix} 0.8781 & 0.1788 \\ 0.0185 & 0.8531 \end{bmatrix}. \tag{14.41c}$$

The two canonical correlations are $r_{C_1} = \sqrt{\lambda_1} = 0.969$ and $r_{C_2} = \sqrt{\lambda_2} = 0.770$. The four canonical vectors are

$$\boldsymbol{a}_1 = [S_{x,x}]^{-1/2}\boldsymbol{e}_1 = \begin{bmatrix} .1788 & -.0522 \\ -.0522 & .0917 \end{bmatrix} \begin{bmatrix} .719 \\ .695 \end{bmatrix} = \begin{bmatrix} .0923 \\ .0263 \end{bmatrix}, \tag{14.42a}$$

$$\boldsymbol{a}_2 = [S_{x,x}]^{-1/2}\boldsymbol{e}_2 = \begin{bmatrix} .1788 & -.0522 \\ -.0522 & .0917 \end{bmatrix} \begin{bmatrix} -.695 \\ .719 \end{bmatrix} = \begin{bmatrix} -.1618 \\ .1022 \end{bmatrix}, \tag{14.42b}$$

$$\boldsymbol{b}_1 = \left[S_{y,y}\right]^{-1/2}\boldsymbol{f}_1 = \begin{bmatrix} .1960 & -.0930 \\ -.0930 & .1699 \end{bmatrix} \begin{bmatrix} .780 \\ .626 \end{bmatrix} = \begin{bmatrix} .0946 \\ .0338 \end{bmatrix}, \tag{14.42c}$$

and

$$\boldsymbol{b}_2 = \left[S_{y,y}\right]^{-1/2}\boldsymbol{f}_2 = \begin{bmatrix} .1960 & -.0930 \\ -.0930 & .1699 \end{bmatrix} \begin{bmatrix} -.626 \\ .780 \end{bmatrix} = \begin{bmatrix} -.1952 \\ .1907 \end{bmatrix}. \tag{14.42d}$$

Alternatively, the eigenvectors $\boldsymbol{e}_m$ and $\boldsymbol{f}_m$ can be obtained through the SVD (Equation 14.35) of the transpose of the matrix in Equation 14.41c (compare the left-hand sides of these two equations). The result is

$$\begin{bmatrix} .8781 & .0185 \\ .1788 & .8531 \end{bmatrix} = \begin{bmatrix} .719 & -.695 \\ .695 & .719 \end{bmatrix} \begin{bmatrix} .969 & 0 \\ 0 & .770 \end{bmatrix} \begin{bmatrix} .780 & .626 \\ -.626 & .780 \end{bmatrix}. \tag{14.43}$$

The canonical correlations are in the diagonal matrix $[R_C]$ in the middle of Equation 14.43. The eigenvectors are in the matrices $[E]$ and $[F]^T$ on either side of it, and can be used to compute the corresponding canonical vectors, as in Equation 14.42. ◇

14.3. MAXIMUM COVARIANCE ANALYSIS (MCA)

14.3.1. Definition of MCA

Maximum covariance analysis (MCA) is a similar technique to CCA, in that it finds pairs of linear combinations of two sets of vector data $\boldsymbol{x}$ and $\boldsymbol{y}$ (Equation 14.2) such that the squares of their covariances

$$\mathrm{Cov}(v_m, w_m) = \boldsymbol{a}_m^T\left[S_{x,y}\right]\boldsymbol{b}_m \tag{14.44}$$

(rather than their correlations, as in CCA) are maximized, subject to the constraint that the vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ are orthonormal. Maximization of squared covariance allows for the possibility that a pair of vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ may yield a negative covariance in Equation 14.44. As in CCA, the number of such pairs $M = \min(I, J)$ is equal to the smaller of the dimensions of the data vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, and each succeeding pair of projection vectors are chosen according to the maximization criterion, subject to the orthonormality constraint. In a typical application to atmospheric data, $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ are both time series of spatial fields, or the leading principal components of these fields, and so their projections in Equation 14.2 form time series also.

Computationally, the vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ are found through a singular value decomposition (Equation 11.72) of the matrix $[S_{x,y}]$ in Equation 14.1, containing the cross-covariances between the elements of $\boldsymbol{x}$ and $\boldsymbol{y}$,

$$\underset{(\mathrm{I}\times\mathrm{J})}{\left[S_{x,y}\right]} = \underset{(\mathrm{I}\times\mathrm{J})}{[A]}\ \underset{(\mathrm{J}\times\mathrm{J})}{[\Omega]}\ \underset{(\mathrm{J}\times\mathrm{J})}{[B]^T}. \tag{14.45}$$

The left singular vectors $\boldsymbol{a}_m$ are the columns of the matrix $[A]$ and the right singular vectors $\boldsymbol{b}_m$ are the columns of the matrix $[B]$ (i.e., the rows of $[B]^{\mathrm{T}}$). The elements ω_m of the diagonal matrix $[\Omega]$ of singular values are the covariances (Equation 14.44) between the pairs of linear combinations in Equation 14.2. Comparison of Equation 14.45 with its CCA counterpart in Equation 14.35 shows that MCA is computed on the basis of the unwhitened (not subjected to Mahalanobis transformations) data vectors $\boldsymbol{x}'$ and $\boldsymbol{y}'$, and that the projection vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ in MCA are not scaled subsequent to the SVD (cf. Equation 14.32 for CCA).

The proportions of the variances of the underlying variables represented by the projections v_m and w_m are

$$R_m^2(\boldsymbol{x}) = \frac{\boldsymbol{a}_m^T [S_{x,x}]\, \boldsymbol{a}_m}{tr([S_{x,x}])} \tag{14.46a}$$

and

$$R_m^2(\boldsymbol{y}) = \frac{\boldsymbol{b}_m^T \left[S_{y,y}\right] \boldsymbol{b}_m}{\mathrm{tr}\left(\left[S_{y,y}\right]\right)}, \tag{14.46b}$$

the numerators of which are $\mathrm{Var}(v_m)$ and $\mathrm{Var}(w_m)$, respectively. The homogeneous correlations are

$$\mathrm{Corr}(v_m, \boldsymbol{x}) = \frac{\boldsymbol{a}_m^T [S_{x,x}][D_x]^{-1}}{\left(\boldsymbol{a}_m^T [S_{x,x}]\, \boldsymbol{a}_m\right)^{1/2}} \tag{14.47a}$$

and

$$\mathrm{Corr}(w_m, \boldsymbol{y}) = \frac{\boldsymbol{b}_m^T \left[S_{y,y}\right] \left[D_y\right]^{-1}}{\left(\boldsymbol{b}_m^T \left[S_{y,y}\right] \boldsymbol{b}_m\right)^{1/2}}, \tag{14.47b}$$

which differ from their counterparts for CCA in Equation 14.7 because the square roots of the variances of the projection variables v_m and w_m in the denominators of Equations 14.47 are not equal to 1. Similarly, the heterogeneous correlations are

$$\mathrm{Corr}(v_m, \boldsymbol{y}) = \frac{\boldsymbol{a}_m^T \left[S_{x,y}\right] \left[D_y\right]^{-1}}{\left(\boldsymbol{a}_m^T [S_{x,x}]\, \boldsymbol{a}_m\right)^{1/2}} \tag{14.48a}$$

and

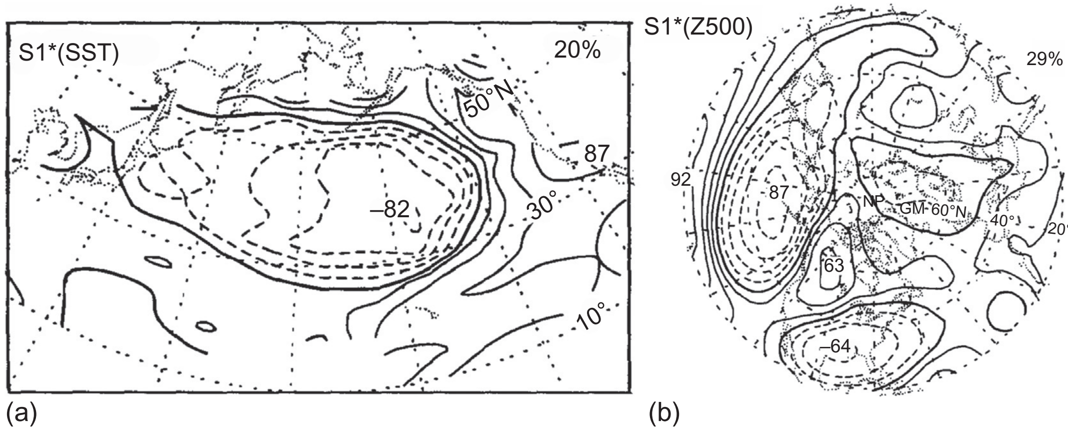

FIGURE 14.5 Homogeneous correlation maps of (a) average winter sea-surface temperatures in the northern Pacific Ocean, and (b) hemispheric winter 500 mb heights, derived from MCA. These are very similar to the corresponding CCA result in Figure 14.1. *From Wallace et al. (1992). © American Meteorological Society. Used with permission.*

$$\mathrm{Corr}(w_m, \boldsymbol{x}) = \frac{\boldsymbol{b}_m^T [S_{y,x}] [D_x]^{-1}}{\left(\boldsymbol{b}_m^T [S_{y,y}] \boldsymbol{b}_m\right)^{1/2}}, \tag{14.48b}$$

which correspond to Equations 14.8.

Because the machinery of the singular value decomposition is used to find the vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$, and the associated covariances ω_m, maximum covariance analysis sometimes unfortunately is known as SVD analysis. As illustrated earlier in this chapter and elsewhere in this book, the singular value decomposition has a rather broader range of uses (e.g., Golub and van Loan, 1996). In recognition of the parallels with CCA, the technique is also sometimes called *canonical covariance analysis* in which case the ω_m are called the canonical covariances. The method is also known as Co-inertia Analysis in the biology literature.

There are two main distinctions between CCA and MCA. The first is that CCA maximizes correlation, whereas MCA maximizes covariance. The leading CCA modes may capture relatively little of the corresponding variances (and thus yield small covariances even if the canonical correlations are high). On the other hand, MCA will find linear combinations with large covariances, which may result more from large variances than a large correlation. The second difference is that the vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ in maximum covariance analysis are orthogonal, and the projections v_m and w_m of the data onto them are in general correlated, whereas the canonical variates in CCA are uncorrelated but the corresponding canonical vectors are not generally orthogonal. Bretherton et al. (1992), Cherry (1996), Tippett et al. (2008), and Van den Dool (2007) compare the two methods in greater detail.

It is not unusual to find similar results for CCA and MCA applied to the same data sets. For example, Figure 14.5 shows a pair of MCA-derived homogeneous correlation patterns for winter northern Pacific SSTs (a) and corresponding 500 mb heights (b), which are both very similar to their counterparts in Figure 14.1 that were based on CCA.

Example 14.4. Maximum Covariance Analysis of the January 1987 Temperature Data

Singular value decomposition of (the transpose of) the cross-covariance submatrix $[S_{x,y}]$ in Equation 14.36c yields

$$\begin{bmatrix} 58.07 & 51.70 \\ 81.63 & 110.8 \end{bmatrix} = \begin{bmatrix} .4876 & .8731 \\ .8731 & -.4876 \end{bmatrix} \begin{bmatrix} 157.4 & 0 \\ 0 & 14.06 \end{bmatrix} \begin{bmatrix} .6325 & .7745 \\ .7745 & -.6325 \end{bmatrix}. \tag{14.49}$$

The results are qualitatively similar to the CCA of the same data in Example 14.1. The first left and right vectors, $\boldsymbol{a}_1 = [0.4876,\ 0.8731]^{\mathrm{T}}$ and $\boldsymbol{b}_1 = [0.6325,\ 0.7745]^{\mathrm{T}}$, respectively, resemble the first pair of canonical vectors $\boldsymbol{a}_1$ and $\boldsymbol{b}_1$ in Example 14.1 in that both put positive weights on both variables in both data sets. Here the weights are closer in magnitude and emphasize the minimum temperatures rather than the maximum temperatures. The covariance between the linear combinations defined by these vectors is 157.4, which is larger than the covariance between any other pair of linear combinations for these data, subject to $||\boldsymbol{a}_1|| = ||\boldsymbol{b}_1|| = 1$. The corresponding correlation is

$$\begin{aligned} \mathrm{Corr}(v_1, w_1) &= \frac{\omega_1}{(\mathrm{Var}(v_1)\,\mathrm{Var}(w_1))^{1/2}} = \frac{\omega_1}{\left(\boldsymbol{a}_1^T [S_{x,x}]\, \boldsymbol{a}_1\right)^{1/2} \left(\boldsymbol{b}_1^T [S_{y,y}]\, \boldsymbol{b}_1\right)^{1/2}}, \\ &= \frac{157.44}{(219.8)^{1/2} (126.3)^{1/2}} = 0.945 \end{aligned} \tag{14.50}$$

which is large, but necessarily smaller than $r_{C_1} = 0.969$ for the CCA of the same data.

The second pair of vectors, $\boldsymbol{a}_2 = [0.8731, -0.4876]^{\mathrm{T}}$ and $\boldsymbol{b}_2 = [0.7745. -0.6325]^{\mathrm{T}}$, are also similar to the second pair of canonical vectors for the CCA in Example 14.1, in that they also describe a contrast between the maximum and minimum temperatures that can be interpreted as being related to the diurnal temperature ranges. The covariance of the second pair of linear combinations is ω_2, corresponding to a correlation of 0.772. This correlation is slightly larger than the second canonical correlation in Example 14.1, but has not been limited by the CCA constraint that the correlations between v_1 and v_2, and w_1 and w_2 must be zero.

The proportions of variability in the original variables captured by the MCA variables are

$$R_1^2(\boldsymbol{x}) = \frac{[.4876\ .8731] \begin{bmatrix} 59.516 & 75.733 \\ 75.433 & 185.467 \end{bmatrix} \begin{bmatrix} .4876 \\ .8731 \end{bmatrix}}{59.516 + 185.467} = 0.897 \tag{14.51a}$$

$$R_2^2(\boldsymbol{x}) = \frac{[.8731 - .4876] \begin{bmatrix} 59.516 & 75.733 \\ 75.433 & 185.467 \end{bmatrix} \begin{bmatrix} .8731 \\ -.4876 \end{bmatrix}}{59.516 + 185.467} = 0.103 \tag{14.52b}$$

$$R_1^2(\boldsymbol{y}) = \frac{[.6325\ .7745] \begin{bmatrix} 61.847 & 56.119 \\ 56.119 & 77.581 \end{bmatrix} \begin{bmatrix} .6325 \\ .7745 \end{bmatrix}}{61.847 + 77.581} = 0.906 \tag{14.53c}$$

and

$$R_2^2(\boldsymbol{y}) = \frac{[.7745 - .6325] \begin{bmatrix} 61.847 & 56.119 \\ 56.119 & 77.581 \end{bmatrix} \begin{bmatrix} .7745 \\ -.6325 \end{bmatrix}}{61.847 + 77.581} = 0.094 \tag{14.54d}$$

◇

The papers of Bretherton et al. (1992) and Wallace et al. (1992) have been influential advocates for the use of maximum covariance analysis. One advantage over CCA that sometimes is cited is that no matrix inversions are required, so that a maximum covariance analysis can be computed even if $n < \max(I, J)$. However, both techniques are subject to similar sampling problems in limited-data situations, so it is not clear that this advantage is of practical importance, and in any case dimension reduction through use of the leading principal components is often employed. Some cautions regarding maximum covariance analysis have been offered by Cherry (1997) and Hu (1997). Newman and Sardeshmukh (1995) emphasize that the $\boldsymbol{a_m}$ and $\boldsymbol{b}_m$ vectors may not represent physical modes of their respective fields, just as the eigenvectors in PCA do not necessarily represent physically meaningful modes.

14.3.2. Forecasting With MCA

The results of a MCA can be used to forecast one of the fields, say $\boldsymbol{y}$, using the $\boldsymbol{x}$ field as the predictor, analogously to the CCA forecasts described in Section 14.2.3. However, since the projections in Equation 14.2 are not uncorrelated for different m in MCA, simultaneous application of M simple linear regressions, as in Equation 14.20 for CCA, will in general not yield optimal predictions. However, if the projection variables in Equation 14.2 have been computed from anomaly vectors $\boldsymbol{x}'$ and $\boldsymbol{y}'$ then the individual MCA prediction regressions will have zero intercept.

Because the v variables for the predictor fields are not independent of the w variables for the predictand fields, in principle any or all M of the v_i may be meaningful predictors for any of the w_j. The resulting M regressions can be represented jointly as

$$\underset{(J\times 1)}{\hat{\boldsymbol{w}}} = \underset{(J\times J)}{[\beta]^T}\,\underset{(J\times 1)}{\boldsymbol{v}} = \underset{(J\times J)}{[\beta]^T}\,\underset{(J\times I)}{[\mathrm{A}]^T}\,\underset{(I\times 1)}{\boldsymbol{x}'}, \tag{14.52}$$

where the jth column of the $(J \times J)$ matrix $[\beta]$ contains the regression coefficients predicting w_j. This equation is the counterpart of Equation 14.23 for CCA except that the parameter matrix $[\beta]$ is in general not diagonal. Wilks (2014b) estimated each column of $[\beta]$ using backward elimination (Section 7.4.2), stopping when nominal p values for all coefficients were no greater than 0.01, although other predictor selection strategies could be used instead.

Equation 14.52 is a multivariate multiple regression (e.g., Johnson and Wichern, 2007), for which the counterpart to Equation 11.37, but including the residuals, is

$$\underset{(J\times n)}{[\hat{W}]} = \underset{(J\times J)}{[\beta]^T}\,\underset{(J\times n)}{[\mathrm{V}]} + \underset{(J\times n)}{[\varepsilon]}. \tag{14.53}$$

Here each of the columns of the matrices $[\hat{W}]$, $[\mathrm{V}]$, and $[\varepsilon]$ contains one of the vectors $\hat{\boldsymbol{w}}$, $\boldsymbol{v}$ and the residual vector $\boldsymbol{\varepsilon}$ for the n training samples. The prediction covariance matrix is then

$$[S_{\hat{\boldsymbol{w}}}] = \frac{1 + \boldsymbol{v}^T\left([\mathrm{V}][\mathrm{V}]^T\right)^{-1}\boldsymbol{v}}{n - M - 1}[\varepsilon][\varepsilon]^T, \tag{14.54}$$

where the residual matrix $[\varepsilon]$ has been computed from the training data using Equation 14.53. Vector mean forecasts for the predictand anomalies $\boldsymbol{y}'$ and their covariance matrix can then be calculated using Equations 14.9 and 14.26 or, if the MCA has been computed using the leading principal components, Equations 14.27 and 14.28.

14.4. REDUNDANCY ANALYSIS (RA)

Both CCA and MCA treat the $\boldsymbol{x}$ and $\boldsymbol{y}$ variables symmetrically, in the sense that the resulting linear combinations and their correlations or covariances are the same if the roles of the two are reversed. In contrast, RA specifically computes the predictand projection vectors $\boldsymbol{b}_m$ in Equation 14.2b which yield maximal variances for the linear combinations w_m conditional on the predictor linear combinations v_m defined by the vectors $\boldsymbol{a_m}$ in Equation 14.2a (e.g., Tippett et al., 2008; Tyler, 1982; Von Storch and Zwiers, 1999). Redundancy analysis is thus specifically oriented toward predicting the $\boldsymbol{y}$ vectors on the basis of the $\boldsymbol{x}$ vectors.

The predictand linear combination vectors $\boldsymbol{b}_m$ are the eigenvectors of the real, symmetric $(J \times J)$ covariance matrix for the predictand vector given the predictors,

$$\left[S_{\hat{y}}\right] = \left[S_{y,x}\right]\left[S_{x,x}\right]^{-1}\left[S_{x,y}\right], \tag{14.55}$$

after which the predictor linear combination vectors $\boldsymbol{a}_m$ can be computed using

$$\boldsymbol{a}_j = \lambda_j^{1/2}\left[S_{x,x}\right]^{-1}\left[S_{x,y}\right]\boldsymbol{b}_j, \tag{14.56}$$

where λ_j is the jth eigenvalue of the matrix in Equation 14.55 and $\boldsymbol{b}_j$ is its corresponding eigenvector.

As will be noted in Section 14.5, the paired pattern finding in RA operates on Mahalanobis-transformed (or whitened) anomaly predictor vectors $\boldsymbol{x}'$, but the predictand anomaly vectors $\boldsymbol{y}'$ are untransformed. Accordingly the predictor linear combinations v_j are uncorrelated with each other so that the regressions predicting the w_j can be considered independently of each other, as in CCA. The forecast vector mean is then computed using Equation 14.52, where the diagonal matrix of regression coefficients is

$$[\beta] = \left([B]^T\left[S_{y,x}\right][A]\right)^{1/2} = [\Lambda]^{1/2}, \tag{14.57}$$

where the columns of the matrices [A] and [B] are the vectors $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$, respectively, and the elements of the diagonal matrix $[\Lambda]$ are the eigenvalues of the matrix in Equation 14.55.

Because the predictor linear combinations v_m are uncorrelated, the prediction covariance matrix is of the same diagonal form as for CCA, Equation 14.24. The individual elements are

$$s^2_{\hat{w}_m} = \left(-\lambda_m + \sum_{k=1}^{J} b^2_{k,m}\right)\left[1 + \frac{1}{n} + \frac{{v_{0,m}}^2}{\sum_{i=1}^{n} {v_{i,m}}^2}\right], \tag{14.58}$$

which differs from its CCA counterpart in Equation 14.25 only in the parenthetical expression for the residual variance. Again the λ_m are the eigenvalues of Equation 14.55. If the predictands are the principal components of the $\boldsymbol{y}$ vectors, the terms in the summation are replaced by $\gamma_m {b^2}_{k,m}$, where the γ_m are the eigenvalues associated with the eigenvectors onto which the predictands are projected.

Relatively little experience with RA for forecasting geophysical fields has yet accumulated, although the forecast skill for seasonal North American temperature forecasts was found to be comparable to those of CCA and MCA in Wilks (2014b).

14.5. UNIFICATION AND GENERALIZATION OF CCA, MCA, AND RA

Swenson (2015) notes that CCA, MCA, and RA are unified mathematically as singular value decompositions of the matrix form

$$[S_{x,x}]^{(\alpha-1)/2}\,[S_{x,y}]\,[S_{y,y}]^{(\beta-1)/2}, \tag{14.59}$$

where α and β are parameters. For $\alpha = \beta = 0$, Equation 14.35 for CCA is obtained. For $\alpha = \beta = 1$, Equation 14.45 for MCA results, because $[S_{x,x}]^0 = [S_{y,y}]^0 = [I]$. Equation 14.59 also encompasses RA, when $\alpha = 0$ and $\beta = 1$, which indicates that the $\boldsymbol{x}$ vectors are subject to the Mahalanobis whitening transformation whereas the $\boldsymbol{y}$ vectors are not.

Intermediate, or *partial whitening*, forms of analysis are also allowed by Equation 14.59, when $0 < \alpha < 1$ and/or $0 < \beta < 1$. Partial whitening partially decorrelates the covariance matrices for $\boldsymbol{x}$ and $\boldsymbol{y}$, so that the off-diagonal elements are reduced in absolute value but not shrunk fully to zero. Partial whitening also yields variances (diagonal elements) that are intermediate between the untransformed elements $s_{x,x}$ or $s_{y,y}$ when $\alpha = 0$ or $\beta = 0$, and unit variance when $\alpha = 1$ or $\beta = 1$. Accordingly the process can be seen as a regularization procedure. The partial whitening of $[S_{x,x}]$ or $[S_{y,y}]$ moves their eigenvalues toward 1 while leaving the eigenvectors unchanged.

Regarding the α and β parameters as free and tunable allows a 2-dimensional continuum of paired relationships to be considered. For example, Figure 14.6 shows contours of cross-validated specification skill in an artificial setting, as functions of α and β. The special cases of CCA, MCA, and RA are indicated by the large circles. For this synthetic example the best cross-validated result is obtained for $\alpha = \beta = 0.35$.

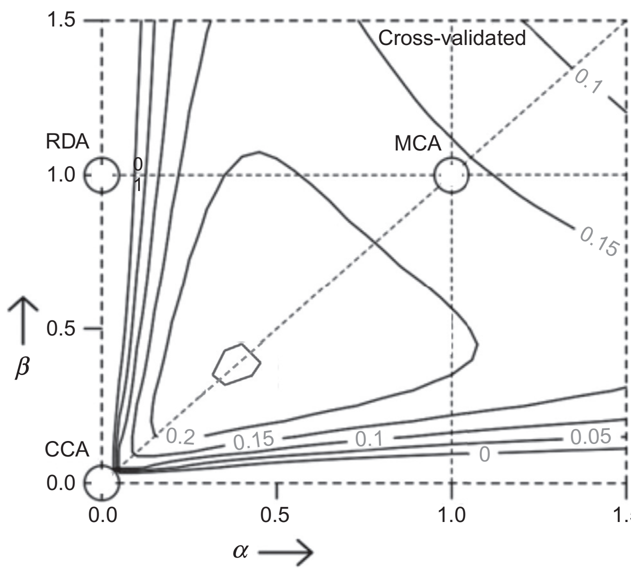

FIGURE 14.6 Dependence of specification skill on the partial-whitening parameters α and β in an artificial setting, showing best performance for $\alpha = \beta = 0.35$ in this case. The special cases of CCA, MCA, and RA are indicated by the circles. *Modified from Swenson (2015). © American Meteorological Society. Used with permission.*

14.6. EXERCISES

14.1. Using the information in Table 14.1 and the data in Table A.1, calculate the values of the canonical variables v_1 and w_1 for 6 January and 7 January.

14.2. The Ithaca maximum and minimum temperatures for 1 January 1988 were $\boldsymbol{x} = (38°\text{F}, 16°\text{F})^T$. Use the CCA in Example 14.1 to "forecast" the Canandaigua temperatures for that day.

14.3. Separate PCAs of the correlation matrices for the Ithaca and Canandaigua data in Table A.1 (after square-root transformation of the precipitation data) yields

$$[E_{\text{Ith}}] = \begin{bmatrix} 0.599 & 0.524 & 0.606 \\ 0.691 & 0.044 & -0.721 \\ 0.404 & -0.851 & 0.336 \end{bmatrix} \text{ and } [E_{\text{Can}}] = \begin{bmatrix} 0.657 & 0.327 & 0.679 \\ 0.688 & 0.107 & -0.718 \\ 0.308 & -0.939 & 0.155 \end{bmatrix}, \quad (14.60)$$

with corresponding eigenvalues $\boldsymbol{\lambda}_{\text{Ith}} = [1.883, 0.927, 0.190]^T$ and $\boldsymbol{\lambda}_{\text{Can}} = (1.904, 0.925, 0.171)^T$. Given also the cross-correlations for these data.

$$[R_{I,C}] = \begin{bmatrix} 0.957 & 0.762 & 0.076 \\ 0.761 & 0.924 & 0.358 \\ 0.166 & 0.431 & 0.904 \end{bmatrix}, \quad (14.61)$$

compute the CCA after truncation to the two leading principal components for each of the locations (and notice that computational simplifications follow from using the principal components), by

a. Computing $[S_C]$, where $\boldsymbol{c}$ is the (4×1) vector $[\boldsymbol{u}_{\text{Ith}}, \boldsymbol{u}_{\text{Can}}]^T$, and then
b. Finding the canonical vectors and canonical correlations.