14 Canonical Correlation Analysis

14.0.0 Overview. Just as EOF analysis (Chapter 13) is used to study the variability of a random vector $\vec{\mathbf{X}}$, *Canonical Correlation Analysis* (CCA) is used to study the correlation structure of a pair of random vectors $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$.

CCA and EOF analyses share similar objectives and similar mathematics. One interpretation of the first EOF $\vec{e}^{\,1}$ of $\vec{\mathbf{X}}$ is that $\vec{\mathbf{X}}^{\mathrm{T}}\vec{e}^{\,1}$ is the linear combination of elements of $\vec{\mathbf{X}}$ with the greatest variance (subject to $\|\vec{e}^{\,1}\| = 1$). The second EOF $\vec{e}^{\,2}$ provides the linear combination $\vec{\mathbf{X}}^{\mathrm{T}}\vec{e}^{\,2}$ with greatest variance that is uncorrelated with $\vec{\mathbf{X}}^{\mathrm{T}}\vec{e}^{\,1}$, and so on. The objective of CCA is to find a pair of patterns $\vec{f}_X^{\,1}$ and $\vec{f}_Y^{\,1}$ (subject to $\|\vec{f}_X^{\,1}\| = \|\vec{f}_Y^{\,1}\| = 1$) so that the *correlation* between linear combinations $\vec{\mathbf{X}}^{\mathrm{T}}\vec{f}_X^{\,1}$ and $\vec{\mathbf{Y}}^{\mathrm{T}}\vec{f}_Y^{\,1}$ is maximized.[1] A second pair of patterns $\vec{f}_X^{\,2}$ and $\vec{f}_Y^{\,2}$ is found so that $\vec{\mathbf{X}}^{\mathrm{T}}\vec{f}_X^{\,2}$ and $\vec{\mathbf{Y}}^{\mathrm{T}}\vec{f}_Y^{\,2}$ are the most strongly correlated linear combinations of $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ that are not correlated with $\vec{\mathbf{X}}^{\mathrm{T}}\vec{f}_X^{\,1}$ and $\vec{\mathbf{Y}}^{\mathrm{T}}\vec{f}_Y^{\,1}$, and so on.

Canonical Correlation Analysis was first described by Hotelling [187].

The 'Canonical Correlation Patterns' of a paired random vector $(\vec{\mathbf{X}}, \vec{\mathbf{Y}})$ are defined in Section 14.1, and their estimation is described in Section 14.2. Examples of some applications are given in Section 14.3. A closely related technique, called *Redundancy Analysis*, is described in Section 14.4.

14.0.1 Introductory Example: Large-scale Temperature and SLP over Europe and Local Weather Elements in Bern. Gyalistras et al. [152] analysed the simultaneous variations of the local climate in Bern (Switzerland) and the troposphere over the North Atlantic in DJF. The state of the local climate in a given season was represented by a 17-dimensional random vector $\vec{\mathbf{X}}$ consisting of the number of days in the season with at least 1 mm of precipitation, and the seasonal means and daily standard deviations of the daily mean, minimum, maximum, and range of temperature, precipitation, wind speed, relative humidity, and relative sunshine duration. The large-scale state of the atmosphere was represented by a vector $\vec{\mathbf{Y}}$ consisting of the near-surface temperature and sea-level pressure (SLP) fields over Europe and the Northeast Atlantic Ocean.

CCA was used to analyse the joint variability of $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$. As noted above, this technique finds pairs of patterns such that the correlation between two corresponding pattern coefficients is maximized.

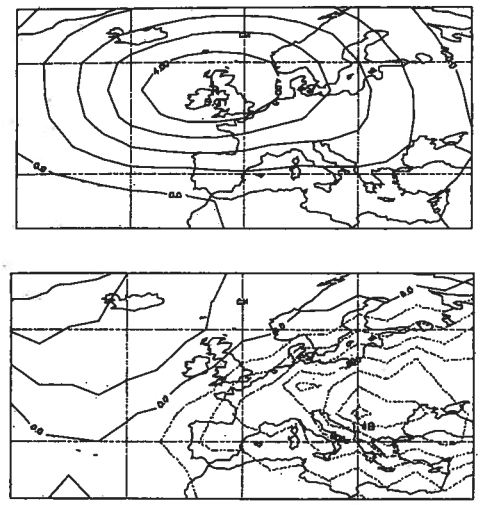
The pair of patterns with the largest correlation is shown in Figure 14.1. The two patterns, one of which consists of two sub-patterns for the pressure and temperature (Figure 14.1, top and middle), have a meaningful physical interpretation. Below normal temperatures in Bern are associated with high pressure over the British Isles and below normal temperatures in the rest of Europe since the correlation between the local climate pattern (bottom panel, Figure 14.1) and the tropospheric pattern (top two panels, Figure 14.1) is negative. Weakened westerly flow is associated with reduced precipitation; the seasonal mean, standard deviation, and number of 'wet' days all tend to be below normal. The large-scale patterns have little effect on wind speed and relative humidity.

The link between the two patterns in Figure 14.1 is strong. The correlation between the coefficient time series (not shown) is -0.89, and the CCA pattern represents a large proportion of the variance of the local climate (Figure 14.2). More than 50% of interannual variance of the seasonal means of daily mean, minimum and maximum temperature are represented by the first CCA pair. They also represent almost 80% of the interannual variance of DJF precipitation and about 75% of the interannual variance of the number of 'wet' days.

14.1 Definition of Canonical Correlation Patterns

14.1.1 One Pair of Patterns. Let us consider an m_X-dimensional random vector $\vec{\mathbf{X}}$ and an

[1] One could also choose $\vec{f}_X^{\,1}$ and $\vec{f}_Y^{\,1}$ to maximize the *covariance* between $\vec{\mathbf{X}}^{\mathrm{T}}\vec{f}_X^{\,1}$ and $\vec{\mathbf{Y}}^{\mathrm{T}}\vec{f}_Y^{\,1}$. Climatologists sometimes call this *SVD analysis* since the patterns are found by obtaining a singular value decomposition of the cross-covariance matrix. See [14.1.7], Bretherton, Smith, and Wallace [64] and Cherry [83].



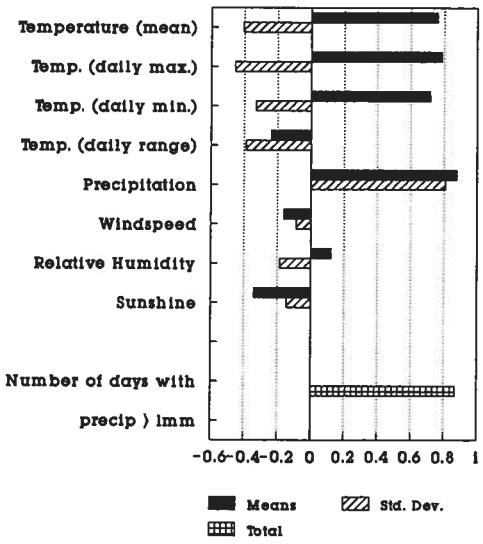


Figure 14.1: *First pair of canonical correlation patterns of $\vec{\mathbf{Y}}$ = (DJF mean SLP, DJF mean temperature) and a vector $\vec{\mathbf{X}}$ of DJF statistics of local weather elements at Bern (Switzerland).*
Top: The SLP part of the first canonical correlation pattern for $\vec{\mathbf{Y}}$.
Middle: The near-surface temperature part of the first canonical correlation pattern for $\vec{\mathbf{Y}}$.
Bottom: The canonical correlation pattern for the local variable $\vec{\mathbf{X}}$.
Note that the correlation between the corresponding pattern coefficients is negative.
From Gyalistras et al. [152].

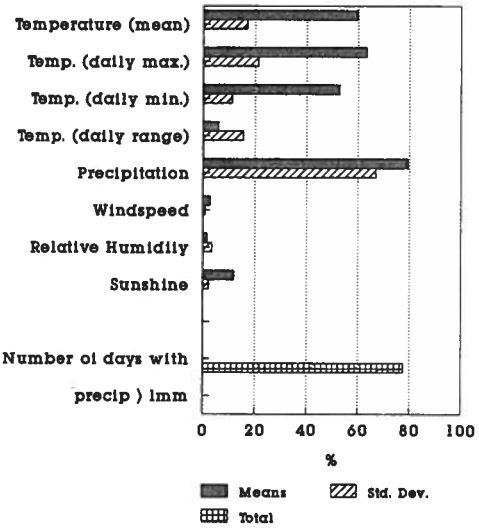


Figure 14.2: *Percentage of year-to-year variance of the local climate variables for Bern represented by the first CCA pair.*

m_Y-dimensional random vector $\vec{\mathbf{Y}}$. We require an m_X-dimensional vector $\vec{f}_X$ and an m_Y-dimensional vector $\vec{f}_Y$ such that the inner products $\beta^X = \langle \vec{\mathbf{X}}, \vec{f}_X \rangle$ and $\beta^Y = \langle \vec{\mathbf{Y}}, \vec{f}_Y \rangle$ have maximum correlation. That is, we want to maximize

$$\rho = \frac{\mathrm{Cov}(\beta^X, \beta^Y)}{\sqrt{\mathrm{Var}(\beta^X)\,\mathrm{Var}(\beta^Y)}} = \frac{\vec{f}_X^{\mathrm{T}} \mathrm{Cov}(\vec{\mathbf{X}}, \vec{\mathbf{Y}})\, \vec{f}_Y}{\sqrt{\mathrm{Var}(\langle \vec{\mathbf{X}}, \vec{f}_X \rangle)\mathrm{Var}(\langle \vec{\mathbf{Y}}, \vec{f}_Y \rangle)}}. \tag{14.1}$$

Note that if a pair of vectors $\vec{f}_X$ and $\vec{f}_Y$ maximizes (14.1), then all vectors $\alpha_X \vec{f}_X$ and $\alpha_Y \vec{f}_Y$ do the same for any nonzero α_X and α_Y. Thus the patterns $\vec{f}_X$ and $\vec{f}_Y$ are subject to arbitrary normalization. In particular, we can choose patterns such that

$$\mathrm{Var}(\langle \vec{\mathbf{X}}, \vec{f}_X \rangle) = \vec{f}_X^{\mathrm{T}} \Sigma_{XX} \vec{f}_X = 1 \tag{14.2}$$

$$\mathrm{Var}(\langle \vec{\mathbf{Y}}, \vec{f}_Y \rangle) = \vec{f}_Y^{\mathrm{T}} \Sigma_{YY} \vec{f}_Y = 1, \tag{14.3}$$

where Σ_{XX} and Σ_{YY} are the covariance matrices of $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$. Then equation (14.1) can be rewritten as

$$\rho = \vec{f}_X^{\mathrm{T}} \Sigma_{XY} \vec{f}_Y, \tag{14.4}$$

where Σ_{XY} is the cross-covariance matrix

$$\Sigma_{XY} = \mathcal{E}\big((\vec{\mathbf{X}} - \vec{\mu}_X)(\vec{\mathbf{Y}} - \vec{\mu}_Y)^{\mathrm{T}}\big).$$

Vectors $\vec{f}_X$ and $\vec{f}_Y$ are found by maximizing

$$\epsilon = \vec{f}_X^{\mathrm{T}}\Sigma_{XY}\vec{f}_Y + \zeta(\vec{f}_X^{\mathrm{T}}\Sigma_{XX}\vec{f}_X - 1) + \eta(\vec{f}_Y^{\mathrm{T}}\Sigma_{YY}\vec{f}_Y - 1), \quad (14.5)$$

where ζ and η are Lagrange multipliers that are used to account for constraints (14.2) and (14.3). Setting the partial derivatives of ϵ to zero, we obtain

$$\frac{\partial\epsilon}{\partial\vec{f}_X} = \Sigma_{XY}\vec{f}_Y + 2\zeta\Sigma_{XX}\vec{f}_X = 0 \quad (14.6)$$

so that

$$\Sigma_{XX}^{-1}\Sigma_{XY}\vec{f}_Y = -2\zeta\vec{f}_X, \quad (14.7)$$

and

$$\frac{\partial\epsilon}{\partial\vec{f}_Y} = \Sigma_{XY}^{\mathrm{T}}\vec{f}_X + 2\eta\Sigma_{YY}\vec{f}_Y = 0, \quad (14.8)$$

which is equivalent to

$$\Sigma_{YY}^{-1}\Sigma_{XY}^{\mathrm{T}}\vec{f}_X = -2\eta\vec{f}_Y. \quad (14.9)$$

Then (14.9) is substituted into (14.7) and vice versa to obtain a pair of eigen-equations for $\vec{f}_X$ and $\vec{f}_Y$:

$$\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{XY}^{\mathrm{T}}\vec{f}_X = 4\zeta\eta\vec{f}_X \quad (14.10)$$

$$\Sigma_{YY}^{-1}\Sigma_{XY}^{\mathrm{T}}\Sigma_{XX}^{-1}\Sigma_{XY}\vec{f}_Y = 4\zeta\eta\vec{f}_Y. \quad (14.11)$$

An argument similar to that used to establish Theorem [13.2.4] proves that the two matrices share the same non-negative eigenvalues.[2] The eigenvectors of the two matrices are related to each other through a simple equation: if $\vec{f}_X$ is a solution of equation (14.10), then $\Sigma_{YY}^{-1}\Sigma_{XY}^{\mathrm{T}}\vec{f}_X$ is a solution of equation (14.11), provided that their joint eigenvalue is nonzero. Finally, equation (14.4) is maximized by letting $\vec{f}_X$ and $\vec{f}_Y$ be the solutions of equations (14.10) and (14.11) that correspond to the largest eigenvalue $\lambda = 4\zeta\eta$.

Now that we have found the canonical random variables $\beta^X = \langle\vec{\mathbf{X}}, \vec{f}_X\rangle$ and $\beta^Y = \langle\vec{\mathbf{Y}}, \vec{f}_Y\rangle$ that are most strongly correlated, the natural next step is to find the value of ρ. Using equations (14.4), (14.6), (14.8), and (14.2), (14.3) in sequence, we find:

$$\begin{aligned}\rho^2 &= \vec{f}_X^{\mathrm{T}}\Sigma_{XY}\vec{f}_Y\vec{f}_Y^{\mathrm{T}}\Sigma_{XY}^{\mathrm{T}}\vec{f}_X \\ &= 4\eta\zeta\,\vec{f}_X^{\mathrm{T}}\Sigma_{XX}\vec{f}_X\vec{f}_Y^{\mathrm{T}}\Sigma_{YY}\vec{f}_Y \\ &= \lambda.\end{aligned}$$

Thus the correlation is the square root of the eigenvalue that corresponds to eigenvectors $\vec{f}_X$ and $\vec{f}_Y$.[3]

14.1.2 More Pairs. The derivation detailed above can now be repeated to obtain $m = \min(m_X, m_Y)$ pairs of patterns $(\vec{f}_X^{\,i}, \vec{f}_Y^{\,i})$ and m corresponding pairs of canonical variates[4]

$$\beta_i^X = \langle\vec{\mathbf{X}}, \vec{f}_X^{\,i}\rangle \quad (14.12)$$

$$\beta_i^Y = \langle\vec{\mathbf{Y}}, \vec{f}_Y^{\,i}\rangle \quad (14.13)$$

with correlation

$$\rho_i = \mathrm{Cov}(\beta_i^X, \beta_i^Y) = \sqrt{\lambda_i}.$$

The patterns and canonical variates are indexed in order of decreasing eigenvalue λ_i. Pairs of canonical variates are uncorrelated. That is, for $i \neq j$,

$$\begin{aligned}\mathrm{Cov}\left(\beta_i^X, \beta_j^X\right) &= \mathrm{Cov}\left(\beta_i^Y, \beta_j^Y\right) \\ &= \mathrm{Cov}\left(\beta_i^X, \beta_j^Y\right) = 0.\end{aligned}$$

14.1.3 The Canonical Correlation Patterns. For simplicity, we assume in this subsection that $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ are of the same dimension m. Then the canonical variates $\vec{\beta}^X = (\beta_1^X, \ldots, \beta_m^X)^{\mathrm{T}}$ and $\vec{\beta}^Y = (\beta_1^Y, \ldots, \beta_m^Y)^{\mathrm{T}}$ can be viewed as the result of coordinate transforms that have been applied to $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$.[5] The transformations relate $\vec{\beta}^X$ and $\vec{\beta}^Y$ to $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ through unknown matrices $\mathcal{F}_X$ and $\mathcal{F}_Y$:

$$\begin{aligned}\vec{\mathbf{X}} &= \mathcal{F}_X\vec{\beta}^X \\ \vec{\mathbf{Y}} &= \mathcal{F}_Y\vec{\beta}^Y.\end{aligned} \quad (14.14)$$

To find $\mathcal{F}_X$, note that

$$\begin{aligned}\vec{\beta}^X &= (\langle\vec{\mathbf{X}}, \vec{f}_X^{\,1}\rangle, \ldots, \langle\vec{\mathbf{X}}, \vec{f}_X^{\,m}\rangle)^{\mathrm{T}} \\ &= \mathbf{f}_X^{\mathrm{T}}\vec{\mathbf{X}}\end{aligned}$$

[2] Note that if $\vec{f}_X$ is a solution of equation (14.10), then $\Sigma_{XX}^{1/2}\vec{f}_X$ is an eigenvector of $(\Sigma_{XX}^{-1/2})^{\mathrm{T}}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{XY}^{\mathrm{T}}\Sigma_{XX}^{-1/2}$. Similarly, $\Sigma_{YY}^{1/2}\vec{f}_Y$ is an eigenvector of $(\Sigma_{YY}^{-1/2})^{\mathrm{T}}\Sigma_{XY}^{\mathrm{T}}\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1/2}$. Since these are non-negative definite matrices, their eigenvalues are real and non-negative.

[3] Note that the sign of the correlation is arbitrary since $\vec{f}_X$ and $\vec{f}_Y$ are determined uniquely only up to their signs.

[4] We assume that $(\Sigma_{XX}^{-1/2})^{\mathrm{T}}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{XY}^{\mathrm{T}}\Sigma_{XX}^{-1/2}$ (or, equivalently, $(\Sigma_{YY}^{-1/2})^{\mathrm{T}}\Sigma_{XY}^{\mathrm{T}}\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1/2}$) has $m = \min(m_X, m_Y)$ distinct, nonzero eigenvalues. Eigenvalues of multiplicity greater than one lead to degeneracy just as in EOF analysis. Uncorrelated canonical variates can still be constructed, but their interpretation is clouded by their non-unique determination. Tools comparable to North's Rule-of-Thumb [13.3.5] are not yet developed for CCA. Note that a pair of degenerate eigenvalues may be an indication of a propagating pattern. See Chapter 15.

[5] The discussion in this subsection is easily generalized to the case in which $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ are not of the same dimension.

where $\mathbf{f}_X$ is the $m \times m$ matrix with eigenvector $\vec{f}_X^{\,i}$ in its ith column. Thus

$$\begin{aligned} \mathrm{Cov}(\vec{\mathbf{X}}, \vec{\beta}^{\,X}) &= \mathrm{Cov}(\vec{\mathbf{X}}, \mathbf{f}_X^{\mathrm{T}}\vec{\mathbf{X}}) \\ &= \mathrm{Cov}(\vec{\mathbf{X}}, \vec{\mathbf{X}})\mathbf{f}_X = \mathbf{\Sigma}_{XX}\mathbf{f}_X. \end{aligned}$$

However, substituting equation (14.14) for $\vec{\mathbf{X}}$, we also have

$$\begin{aligned} \mathrm{Cov}(\vec{\mathbf{X}}, \vec{\beta}^{\,X}) &= \mathrm{Cov}(\mathcal{F}_X\vec{\beta}^{\,X}, \vec{\beta}^{\,X}) \\ &= \mathcal{F}_X\,\mathrm{Cov}(\vec{\beta}^{\,X}, \vec{\beta}^{\,X}) = \mathcal{F}_X \end{aligned}$$

since $\mathrm{Cov}(\vec{\beta}^{\,X}, \vec{\beta}^{\,X}) = \mathcal{I}$. Thus

$$\mathcal{F}_X = \mathbf{\Sigma}_{XX}\mathbf{f}_X \tag{14.15}$$

and similarly

$$\mathcal{F}_Y = \mathbf{\Sigma}_{YY}\mathbf{f}_Y. \tag{14.16}$$

The columns of $\mathcal{F}_X$ and $\mathcal{F}_Y$, $\vec{F}_X^i$ and $\vec{F}_Y^i$, are called the *canonical correlation patterns*.[6] The canonical variates β_i^X and β_i^Y are also often called *canonical correlation coordinates*. Since the canonical correlation coordinates are normalized to unit variance, the canonical correlation patterns are expressed in the units of the field they represent, and they indicate the 'typical' strength of the mode of covariation described by the patterns.

While the matrix-vector representations of $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ in (14.14) are convenient for the derivation of $\mathcal{F}_X$ and $\mathcal{F}_Y$, they are not very evocative. Therefore, note that (14.14) can also be written as

$$\begin{aligned} \vec{\mathbf{X}} &= \textstyle\sum_i \beta_i^X \vec{F}_X^i \\ \vec{\mathbf{Y}} &= \textstyle\sum_i \beta_i^Y \vec{F}_Y^i. \end{aligned} \tag{14.17}$$

This allows us to see more clearly that (14.14) describes an expansion of $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ with respect to their corresponding canonical correlation patterns. It also suggests that it may be possible to approximate $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ by truncating the summation in (14.17).

14.1.4 Computational Aspects. Once we know one set of vectors, say $\vec{f}_X^{\,i}$, all other vectors are easily obtained through simple matrix operations. Let us assume that we have the vectors $\vec{f}_X^{\,i}$. Then (14.15) yields $\vec{F}_X^i$. In [14.1.1] we noted that $\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{XY}^{\mathrm{T}}\vec{f}_X^{\,i}$ is equal to $\vec{f}_Y^{\,i}$ after suitable normalization. Application of (14.16) gives the last set of vectors, the $\vec{\mathbf{Y}}$-canonical correlation patterns $\vec{F}_Y^i$. It is therefore necessary to solve only the smaller of the two eigenproblems (14.10) and (14.11).

14.1.5 Coordinate Transformations. What happens to the canonical correlation patterns and correlations when coordinates are transformed by an invertible matrix $\mathcal{L}$ through $\mathcal{L}\vec{\mathbf{X}} = \vec{\mathbf{Z}}$? For simplicity we assume random vector $\vec{\mathbf{Y}}$ is unchanged.

To get the same maximum correlation (14.1), we have to transform the patterns $\vec{f}_X^{\,i}$ with $\mathcal{L}^{-1}$,

$$\vec{f}_Z^{\,i} = (\mathcal{L}^{-1})^{\mathrm{T}}\vec{f}_X^{\,i}. \tag{14.18}$$

Thus the canonical correlation coordinates $\beta_i^X = \langle \vec{f}_Z^{\,i}, \vec{\mathbf{Z}} \rangle = \langle \vec{f}_X^{\,i}, \vec{\mathbf{X}} \rangle$ are unaffected by the transformation. Note that relation (14.18) can also be obtained by verifying that $\vec{f}_Z^{\,i}$ and $\vec{f}_X^{\,i}$ are eigenvectors of the CCA matrices $\mathbf{\Sigma}_{ZZ}^{-1}\mathbf{\Sigma}_{ZY}\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{ZY}^{\mathrm{T}}$ and $\mathbf{\Sigma}_{XX}^{-1}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{XY}^{\mathrm{T}}$ with the same eigenvalues.

The canonical correlation patterns $\vec{F}_X^i$ are determined by the covariance matrix of $\vec{\mathbf{X}}$ and the $\vec{f}_X^{\,i}$-pattern (14.15). Therefore,

$$\begin{aligned} \vec{F}_Z^i &= \mathbf{\Sigma}_{ZZ}\vec{f}_Z^{\,i} \\ &= \mathcal{L}\mathbf{\Sigma}_{XX}\mathcal{L}^{\mathrm{T}}(\mathcal{L}^{-1})^{\mathrm{T}}\vec{f}_X^{\,i} \\ &= \mathcal{L}\mathbf{\Sigma}_{XX}\vec{f}_X^{\,i} \\ &= \mathcal{L}\vec{F}_X^i. \end{aligned} \tag{14.19}$$

Thus the canonical correlation patterns are transformed in the same way as the random vector $\vec{\mathbf{X}}$. We may conclude that *the CCA is invariant under coordinate transformations.*

14.1.6 CCA after a Transformation to EOF Coordinates. The CCA algebra becomes considerably simpler if the data are transformed into EOF space before the analysis (Barnett and Preisendorfer [21]). Suppose that only the first k_X and k_Y EOFs are retained, so that

$$\begin{aligned} \vec{\mathbf{X}} &\approx \textstyle\sum_i^{k_X} \alpha_i^{X+}\vec{e}_X^{\,i+} \\ \vec{\mathbf{Y}} &\approx \textstyle\sum_i^{k_Y} \alpha_i^{Y+}\vec{e}_Y^{\,i+}, \end{aligned} \tag{14.20}$$

where we have used the renormalized versions (13.20, 13.21) of the EOFs and their coefficients $\alpha_i^+ = (\lambda_i)^{-1/2}\alpha_i$ and $\vec{e}^{\,i+} = (\lambda_i)^{1/2}\vec{e}^{\,i}$. The CCA is then applied to the random vectors $\vec{\mathbf{X}}' = (\alpha_1^{X+}, \ldots, \alpha_{k_X}^{X+})^{\mathrm{T}}$ and $\vec{\mathbf{Y}}' = (\alpha_1^{Y+}, \ldots, \alpha_{k_Y}^{Y+})^{\mathrm{T}}$.

An advantage of this approach is that it is often possible to use only the first few EOFs.

[6] Note that neither the eigenvectors $\vec{f}_X^{\,i}$ and $\vec{f}_Y^{\,i}$ nor the canonical correlation patterns $\vec{F}_X^i$ and $\vec{F}_Y^i$ are generally orthogonal. However, the columns of $\mathbf{\Sigma}_{XX}^{1/2}\mathbf{f}_X = \mathbf{\Sigma}_{XX}^{-1/2}\mathcal{F}_X$ and $\mathbf{\Sigma}_{YY}^{1/2}\mathbf{f}_Y = \mathbf{\Sigma}_{YY}^{-1/2}\mathcal{F}_Y$ are orthonormal.

Discarding the high-index EOFs can reduce the amount of noise in the problem by eliminating poorly organized, small-scale features of the fields involved.

Another advantage is that the algebra of the problem is simplified since $\mathbf{\Sigma}_{X'X'}$ and $\mathbf{\Sigma}_{Y'Y'}$ are both identity matrices. Thus, according to equations (14.10, 14.11), $\vec{f}_{X'}^{\,i}$ and $\vec{f}_{Y'}^{\,i}$ are eigenvectors of $\mathbf{\Sigma}_{X'Y'}\mathbf{\Sigma}_{X'Y'}^{\mathrm{T}}$ and $\mathbf{\Sigma}_{X'Y'}^{\mathrm{T}}\mathbf{\Sigma}_{X'Y'}$ respectively. Since these are non-negative definite symmetric matrices, the eigenvectors are orthogonal. Moreover, the canonical correlation patterns $\vec{F}_{X'}^{\,i} = \vec{f}_{X'}^{\,i}$ and $\vec{F}_{Y'}^{\,i} = \vec{f}_{Y'}^{\,i}$.

A minor disadvantage is that the patterns are given in the coordinates of the re-normalized EOF space (14.20). To express the pattern in the original coordinate space it is necessary to reverse transformation (14.20) with (14.18) and (14.19):

$$\begin{aligned}
\vec{f}_X^{\,i} &= \textstyle\sum_{j=1}^{k_X} (\lambda_j^X)^{1/2} (\vec{f}_{X'}^{\,i})_j \vec{e}_X^{\,j} \\
\vec{f}_Y^{\,i} &= \textstyle\sum_{j=1}^{k_Y} (\lambda_j^Y)^{1/2} (\vec{f}_{Y'}^{\,i})_j \vec{e}_Y^{\,j} \\
\vec{F}_X^{\,i} &= \textstyle\sum_{j=1}^{k_X} (\lambda_j^X)^{-1/2} (\vec{f}_{X'}^{\,i})_j \vec{e}_X^{\,j} \\
\vec{F}_Y^{\,i} &= \textstyle\sum_{j=1}^{k_Y} (\lambda_j^Y)^{-1/2} (\vec{f}_{Y'}^{\,i})_j \vec{e}_Y^{\,j}
\end{aligned} \tag{14.21}$$

where $(\cdot)_j$ denotes the jth element of the vector contained within the brackets. The canonical correlation patterns are no longer orthogonal after this backtransformation, and vectors $\vec{f}^{\,i}$ and $\vec{F}^i$ are no longer identical.[7]

14.1.7 Maximizing Covariance—the 'SVD Approach.' Another way to identify pairs of coupled patterns $\vec{p}_X^{\,i}$ and $\vec{p}_Y^{\,i}$ in random fields $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ is to search for orthonormal sets of vectors such that the covariance between the expansion coefficients $\alpha_i^X = \langle \vec{\mathbf{X}}, \vec{p}_X^{\,i} \rangle$ and $\alpha_i^Y = \langle \vec{\mathbf{Y}}, \vec{p}_Y^{\,i} \rangle$,

$$\mathrm{Cov}(\alpha_i^X, \alpha_i^Y) = (\vec{p}_X^{\,i})^{\mathrm{T}} \mathbf{\Sigma}_{XY} \vec{p}_Y^{\,i}, \tag{14.22}$$

is maximized. Note that we explicitly require orthonormal vectors so that $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ can be expanded as $\vec{\mathbf{X}} = \sum \alpha_i^X \vec{p}_X^{\,i}$ and $\vec{\mathbf{Y}} = \sum \alpha_i^Y \vec{p}_Y^{\,i}$. The solution of (14.22) is obtained as in [14.1.1] by using Lagrange multipliers to enforce the constraints $(\vec{p}_X^{\,i})^{\mathrm{T}} \vec{p}_X^{\,i} = 1$ and $(\vec{p}_Y^{\,i})^{\mathrm{T}} \vec{p}_Y^{\,i} = 1$. The result is a system of equations,

$$\begin{aligned}
\mathbf{\Sigma}_{XY} \vec{p}_Y^{\,i} &= s_X \vec{p}_X^{\,i} \\
\mathbf{\Sigma}_{XY}^{\mathrm{T}} \vec{p}_X^{\,i} &= s_Y \vec{p}_Y^{\,i},
\end{aligned} \tag{14.23}$$

that can be solved by a singular value decomposition (Appendix B). The same solution is obtained by substituting the two equations into each other to obtain

$$\begin{aligned}
\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_{XY}^{\mathrm{T}} \vec{p}_X^{\,i} &= \lambda_i \vec{p}_X^{\,i} \\
\mathbf{\Sigma}_{XY}^{\mathrm{T}}\mathbf{\Sigma}_{XY} \vec{p}_Y^{\,i} &= \lambda_i \vec{p}_Y^{\,i},
\end{aligned}$$

where $\lambda_i = s_X s_Y$. These equations share the same eigenvalues $\lambda_i > 0$, and their normalized eigenvectors are related by

$$\begin{aligned}
\vec{p}_Y^{\,i} &= \frac{\mathbf{\Sigma}_{XY}^{\mathrm{T}} \vec{p}_X^{\,i}}{\|\mathbf{\Sigma}_{XY}^{\mathrm{T}} \vec{p}_X^{\,i}\|} \\
\vec{p}_X^{\,i} &= \frac{\mathbf{\Sigma}_{XY}^{\mathrm{T}} \vec{p}_Y^{\,i}}{\|\mathbf{\Sigma}_{XY}^{\mathrm{T}} \vec{p}_Y^{\,i}\|}.
\end{aligned}$$

It is easily shown that $\mathrm{Cov}(\alpha_i^X, \alpha_i^Y) = \lambda_i^{1/2}$. Thus the pair of patterns associated with the largest eigenvalue maximizes the covariance. The pair of patterns associated with the second largest eigenvalue and orthogonal to the first pair maximize the covariability that remains in $\vec{\mathbf{X}} - \alpha_1^X \vec{p}_X^{\,1}$ and $\vec{\mathbf{Y}} - \alpha_1^Y \vec{p}_Y^{\,1}$, and so on.

This method is often called 'SVD' analysis. This wording is misleading because it mixes the definition of a statistical parameter with the algorithm used to calculate the parameter. These patterns *can* be calculated by SVD but there are other ways, such as conventional eigen-analysis, to get the same information. Patterns $\vec{p}_X^{\,i}$ and $\vec{p}_Y^{\,i}$ are often called *left* and *right singular vectors*. The nomenclature is again misleading because the relevant property of these vectors is that they maximize covariance. We therefore call this method *Maximum Covariance Analysis* (MCA) and call the vectors *Maximum Covariance Patterns*.

Two properties of MCA are worth mentioning.

- MCA is invariant under coordinate transformation only if the transformation is orthogonal. The eigenvalues, and thus the degree of covariability, change when the transformation is non-orthonormal.
- MCA coefficients α_i^X and α_j^X, $i \neq j$, are generally correlated. They are uncorrelated when $\mathbf{\Sigma}_{XX} = \sigma_X^2 \mathcal{I}$. This also applies to $\vec{\mathbf{Y}}$-coefficients.

See Wallace, Smith, and Bretherton [411] for examples.

14.1.8 Principal Prediction Patterns. Suppose $\{\vec{\mathbf{Z}}_t\}$ is a multivariate time series and define

$$\vec{\mathbf{X}}_t = \vec{\mathbf{Z}}_t \text{ and } \vec{\mathbf{Y}}_t = \vec{\mathbf{Z}}_{t+\tau} \tag{14.24}$$

[7] Note the similarity between this discussion and that in [14.1.4].

for some positive lag τ. Application of the CCA algorithm, with prior EOF truncation if the dimension of $\vec{\mathbf{Z}}_t$ is large, identifies patterns $\vec{F}_0^i = \vec{F}_X^i$ and $\vec{F}_\tau^i = \vec{F}_Y^i$ that tend to appear together, that is, patterns with a fixed time lag in the same variable. Thus the presence of $\vec{F}_0^i$ at a given time indicates that it is likely that pattern $\vec{F}_\tau^i$ will emerge τ time units later. Because of the properties of CCA, patterns $\vec{F}_0^i$ and $\vec{F}_\tau^i$ depict the present and future parts of $\vec{\mathbf{Z}}_t$ that are most strongly related. In other words, they are the best linearly auto-predictable components in $\vec{\mathbf{Z}}_t$.[8] An example is given in [14.3.7].

14.2 Estimating Canonical Correlation Patterns

14.2.1 Estimation. Estimates of canonical correlation patterns and coefficients are obtained in the obvious way by replacing $\mathbf{\Sigma}_{XX}$, $\mathbf{\Sigma}_{YY}$, and $\mathbf{\Sigma}_{XY}$ with corresponding estimates. We recommend that the problem be kept small by approximating the data with truncated EOF expansions (see [14.1.5] and also Bretherton et al. [64]). This has the added benefit of eliminating small-scale spatial noise.

14.2.2 Making Inferences. As noted previously, very little is known about the sampling variability of the eigenvectors or canonical correlation patterns. However, there are some useful asymptotic results for making inferences about the canonical correlations themselves.

Bartlett [32] proposed a test of the null hypothesis H_0: $\rho_{l+1} = \cdots = \rho_m = 0$ that the last $m - l$ canonical correlations are zero when it is *known* that the first l are nonzero. Here $m = \min(m_X, m_Y)$. Bartlett's test can be used when the canonical correlations have been estimated from a sample $\{(\vec{\mathbf{x}}_1, \vec{\mathbf{y}}_1), \ldots, (\vec{\mathbf{x}}_n, \vec{\mathbf{y}}_n)\}$ of independent realizations of random vectors $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ that are jointly multivariate normal. The test statistic (Bartlett [32])

$$\chi^2 = -(n - 1 - l - \frac{1}{2}(m_X + m_Y + 1)) + \sum_{i=1}^{l} \widehat{\rho}_i^{-2} \ln\Big(\prod_{i=l+1}^{m} (1 - \widehat{\rho}_i^2)\Big), \quad (14.25)$$

where $\widehat{\rho}_i = \sqrt{\hat{\lambda}_i}$, is approximately distributed as $\chi^2((m_X - l)(m_Y - l))$ under H_0. The test is performed at the $(1 - \tilde{p}) \times 100\%$ significance level by comparing χ^2 (14.25) against the $\tilde{p}$-quantile of the approximating χ^2 distribution (see Appendix E).

Glynn and Muirhead [142] give a bias correction for $\widehat{\rho}_i$ and also give an expression for the asymptotic variance of the corrected estimator that is useful for constructing confidence intervals. Using the Fisher z-transform (recall [8.2.3]), Glynn and Muirhead show that if

$$\theta_i = \frac{1}{2}\ln\Big(\frac{1 + \rho_i}{1 - \rho_i}\Big) \text{ and } z_i = \frac{1}{2}\ln\Big(\frac{1 + \widehat{\rho}_i}{1 - \widehat{\rho}_i}\Big),$$

then the bias of

$$\widehat{\theta}_i = z_i - \frac{1}{2n\widehat{\rho}_i}\Big(m_X + m_Y - 2 + \widehat{\rho}_i^2 + 2(1 - \widehat{\rho}_i^2) \sum_{\substack{j=1; \\ j \neq i}}^{m} \frac{\widehat{\rho}_j^2}{\widehat{\rho}_i^2 - \widehat{\rho}_j^2}\Big)$$

is approximately $\mathcal{O}(n^{-2})$ and

$$\mathrm{Var}(\widehat{\theta}_i) = \frac{1}{n} + \mathcal{O}(n^{-2}).$$

Thus the bounds for an approximate $\tilde{p} \times 100\%$ confidence interval for ρ_i are given by

$$\tanh(\widehat{\theta}_i \pm z_{(1+\tilde{p})/2}/\sqrt{n}), \quad (14.26)$$

where $z_{(1+\tilde{p})/2}$ is the $(1 + \tilde{p})/2$-quantile of the standard normal distribution (Appendix D). Muirhead and Waternaux [282] show that asymptotic statistics like equations (14.25, 14.26) are not particularly robust against departures from the multivariate normal assumption. Use of the bootstrap (see Section 5.5) is probably the best practical alternative when this is a concern.

One question rarely mentioned in the context of CCA is the size of sample needed to make good estimates and inferences. Thorndike [365, pp. 183–184] suggests that $n > 10(m_X + m_Y) + 50$ is a reasonable rule of thumb, and argues that $n > (m_X + m_Y)^2 + 50$ may be needed for some purposes. Our experience, however, is that much smaller samples can provide meaningful information about the first few patterns and correlations. However, be aware that the asymptotic results discussed above are not likely to hold under these circumstances. The Monte Carlo experiments discussed in the next subsection give some further insight into what can be accomplished with small samples.

[8] It seems that the idea was first suggested by Hasselmann in an unpublished paper in 1983 but it was not pursued until 1996 [103].

mode		$k = 20$					$n = 250$		
i	ρ^i_{xy}	$n =$	50	100	500	1000	$k =$	10	30	50
1	0.69		0.96	0.83	0.70	0.69		0.68	0.71	0.74
2	0.60		0.92	0.76	0.59	0.58		0.58	0.61	0.65
3	0.37		0.79	0.51	0.33	0.31		0.30	0.36	0.43
4	0.11		0.54	0.28	0.10	0.09		0.06	0.16	0.27
5	0.07		0.46	0.23	0.08	0.06		0.03	0.13	0.25

Table 14.1: *The means of 100 canonical correlation estimates computed from simulated samples of n pairs of 251-dimensional random fields (see text). For brevity, only five of the 10 canonical correlations are listed. The true correlations ρ^i_{xy} are given in the second column; the results obtained for variable time series lengths n, with an EOF truncation of $k = 20$, are given in columns three to six. The effect of including different numbers of EOFs k, using a fixed time series length of $n = 250$, is listed in columns seven to nine. From Borgert [55].*

14.2.3 Monte Carlo Experiments. Borgert [55] conducted a Monte Carlo study of the performance of CCA on EOF truncated [14.1.6] data. He simulated a pair of 251-point random fields $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ that consisted of a random linear combination of 10 pairs of patterns. Each pair of patterns was multiplied by a pair of random coefficients that were independent of all other pairs of coefficients. Thus the random coefficients are the true canonical variables. Each pair of random coefficients was generated from a different bivariate auto-regressive process. In this way the cross-correlations between the pairs of canonical variates, the true canonical correlations, were known. Thus Borgert was able to simulate a pair of random fields with known canonical correlations and patterns.

Borgert used this tool to generate 100 independent samples for a number of combinations of sample size n and EOF truncation point $k = k_X = k_Y$. A canonical correlation analysis was performed on each sample, and statistics assessing the average quality of the CCA were gathered for each combination of n and k. He found that the CCA was really able to identify the correct pairs of patterns: the estimated patterns were close to the prescribed patterns. However, as exemplified in Table 14.1, there were considerable biases in the estimated correlations if too many EOFs were retained or if the time series were too short.

Bretherton et al. [64] reviewed a number of techniques for diagnosing coupled patterns and intercomparing them in a series of small Monte Carlo experiments. They found that CCA with *a priori* EOF truncation and Maximum Covariance Analysis were more robust than the other techniques considered.

14.2.4 Irregularly Distributed Gaps in the Data. One way to cope with missing data is to fill the gaps by spatial or temporal interpolation. However, this is unsatisfactory if more than just a small amount of data is missing because we end up trying to diagnose connections between real data on the one hand and imputed data with much lower information content on the other. A better procedure is to use only the data that are actually available. This can be achieved by the procedure already outlined in [13.2.7]. The various matrices, such as Σ_{XX}, are estimated by forming sums over only the available pairs of observations (13.31):

$$\widehat{\sigma}_{ij} = \frac{1}{|K_i \cap K_j|} \sum_{k \in K_i \cap K_j} (\mathbf{x}_{ki} - \widehat{\mu}_i)(\mathbf{x}_{kj} - \widehat{\mu}_j)^*$$

where $K_i = \{k\text{: component } i \text{ of } \vec{\mathbf{x}}_k \text{ is not missing}\}$, the notation $|\cdot|$ indicates the number of elements in a set, and $\widehat{\mu}_i = \frac{1}{|K_i|} \sum_{k \in K_i} \mathbf{x}_{ki}$. As with EOFs, the calculation of the time coefficients can no longer be done by means of the dot products (14.12) and (14.13). Instead coefficients are determined by least squares, as in equation (13.32).

14.3 Examples

14.3.0 Overview. We will present three examples in this section. The joint variability of a pair of large-scale fields is examined for evidence of a cause-and-effect relationship between the occurrence of large-scale sea-level air pressure and sea-surface temperatures anomalies in the North

Atlantic [14.3.1,2]. In the second example, one of the vector times series is again North Atlantic sea-level pressure but the second 'partner' in the CCA is a regional scale variable, namely, precipitation on the Iberian Peninsula [14.3.3,4]. This example is used to demonstrate *statistical downscaling* of GCM output. The last example [14.3.5] illustrates the *Principal Prediction Patterns* introduced in [14.1.7].

The literature also contains many other examples of applications of CCA. Bretherton et al. [64] cite several studies, including classic papers by Barnett and Preisendorfer [21], Nicholls [293] and Barnston and colleagues [26, 28, 30, 346].

14.3.1 North Atlantic SLP and SST: Data and Results. CCA is used to analyse the relationship between $\vec{X}$ = *monthly mean sea-level pressure* (SLP) and $\vec{Y}$ = *sea-surface temperature* (SST) over the North Atlantic in northern winter (DJF) (see Zorita et al. [438] for details). The data are time series of monthly means of SLP and SST on a grid over the North Atlantic north of about 20° N, for DJF of 1950 to 1986. Anomalies were obtained at each grid point by subtracting the long-term monthly mean from the original values.

The coefficients of the first five EOFs of both fields were retained for the subsequent CCA. They represent 87% and 62% of the total variance or SLP and SST respectively. To check the sensitivity of the results to EOF truncation, the same calculations were performed using five SLP EOFs and either 10 or 15 SST EOFs (77% and 84%, respectively) and essentially the same results were obtained.

The CCA yields two pairs of patterns that describe the coherent variations of the SST and SLP fields. The two patterns are dominant in describing SLP and SST variance.

The first pair of patterns, $\vec{F}^1_{SLP}$ and $\vec{F}^1_{SST}$, which corresponds to a canonical correlation of 0.56, represents 21% of the variance of monthly mean SLP and 19% of the variance of monthly mean SST (Figure 1.13).[9] The two patterns are consistent with the hypothesis first suggested by Bjerknes that atmospheric anomalies cause SST anomalies. The main features of the SLP pattern are a decrease of the westerly wind at about 50° N, and an anomalous cyclonic circulation centred at 40° W and 30° N.[10] North of the cyclone, the ocean surface is warmer than normal when the westerly wind is reduced. West of the cyclone, just downstream from the cold American continent, the ocean is substantially cooled. The SST anomalies off the African coast are a local response to anomalous winds; coastal upwelling is reduced when there are weaker than normal northerly winds. In contrast, when the circulation produces enhanced westerlies and anomalous anticyclonic flow in the southern part of the area, opposite SST anomalies are expected. The canonical correlation coefficient time series also support the Bjerknes hypothesis: the one month lag correlation is 0.65 when SLP leads SST but it is only 0.09 if SLP lags.

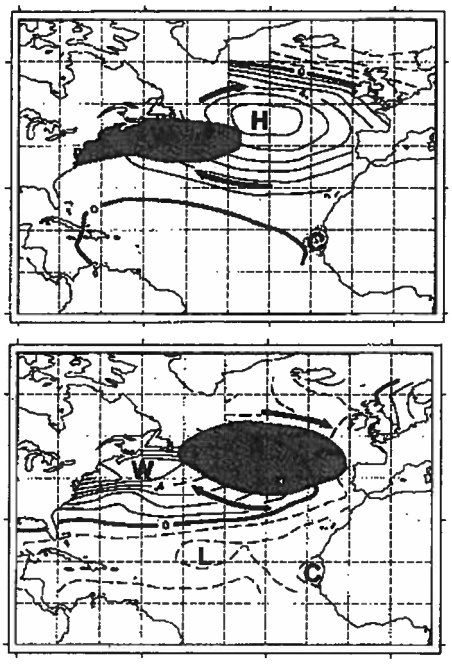


Figure 14.3: *The second pair of canonical patterns for monthly mean SLP and SST over the North Atlantic in DJF. The dark shading on each pattern identifies the main positive feature of the opposing pattern.*
Top: SLP, contour interval: 1 hPa,
Bottom: SST, contour interval: 0.1 K.
From Zorita et al. [438].

The coefficients of the second pair of patterns, $\vec{F}^2_{SLP}$ and $\vec{F}^2_{SST}$, have correlation 0.47 (Figure 14.3). The SLP pattern represents 31% of the total variance and is similar to the first SLP EOF (Figure 13.8), which is related to the North Atlantic Oscillation (see also [13.5.5] and

[9] We have dropped the '$\hat{\ }$' notation for now, but be aware that the patterns are parameter estimates. The same applies to canonical coordinate time series when they are discussed.

[10] We use the geostrophic wind relationship for the derivation of approximate wind anomalies from pressure anomalies.

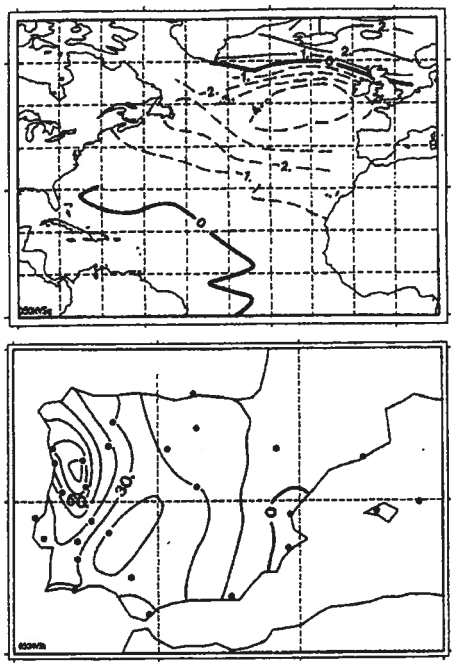


Figure 14.4: *First pair of canonical correlation patterns of the North Atlantic winter mean sea-level pressure* $\vec{\mathbf{Y}}$ *and a vector* $\vec{\mathbf{X}}$ *of seasonal means of precipitation at a number of Iberian locations [403].*

Figure 13.6). The structure of this pair of patterns is also consistent with the Bjerknes hypothesis. The one month lag correlation is 0.48 when SLP leads and 0.03 when SLP lags.

14.3.2 North Atlantic SLP and SST: Discussion. We described conventional and rotated EOF analysis of the same data in [13.5.6,7]. The CCA of SLP and SST suggests why rotation had a marked effect on the SST EOFs but not on the SLP EOFs. The coherent variations in the atmosphere (SLP) are caused by large-scale internal atmospheric processes so that the EOFs have a simple large-scale structure. In case of the ocean (SST), however, the coherent variations (EOFs) are the oceanic response to the large-scale atmospheric variations. This response really does not have simple structure (recall our description in [14.3.3] of the ocean's response to NAO variations).

14.3.3 North Atlantic SLP and Iberian Rainfall: Analysis and Historic Reconstruction. In this example, winter (DJF) mean precipitation from a number of rain gauges on the Iberian Peninsula is related to the air-pressure field over the North Atlantic (see [403] for details). CCA was used to obtain a pair of canonical correlation pattern estimates $\vec{F}^1_{SLP}$ and $\vec{F}^1_{pre}$ (Figure 14.4), and corresponding time series $\beta_1^{SLP}(t)$ and $\beta_1^{pre}(t)$ of canonical variate estimates. These strongly correlated modes of variation (the estimated canonical correlation is 0.75) represent about 65% and 40% of the total variability of seasonal mean SLP and Iberian Peninsula precipitation respectively. The two patterns represent a simple physical mechanism: when $\vec{F}^1_{SLP}$ has a strong positive coefficient, enhanced cyclonic circulation advects more maritime air onto the Iberian Peninsula so that precipitation in the mountainous northwest region ($\vec{F}^1_{pre}$) is increased.

Since the canonical correlation is large, the results of the CCA can be used to forecast or specify winter mean precipitation on the Iberian peninsula from North Atlantic SLP. The first step is to connect $\beta_1^{pre}(t)$ and $\beta_1^{SLP}(t)$ with a simple linear model $\beta_1^{pre}(t) = a\beta_1^{SLP}(t) + \epsilon$. Since $\beta_1^{pre}(t)$ and $\beta_1^{SLP}(t)$ are normalized to unit variance, the least squares estimate of coefficient a is the canonical correlation ρ_1. Given a realization of $\beta_1^{SLP}(t)$, the canonical variate for precipitation can be forecast as $\widehat{\beta}_1^{pre}(t) = \rho_1\beta_1^{SLP}(t)$, and thus the precipitation field is forecast as

$$\widehat{\vec{\mathbf{R}}} = \widehat{\beta}_1^{pre}(t)\vec{F}^1_{pre} = \rho_1\beta_1^{SLP}(t)\vec{F}^1_{pre}. \qquad (14.27)$$

Similarly, if several useful canonical correlation patterns had been found, Iberian winter mean precipitation could be forecast or specified as

$$\widehat{\vec{\mathbf{R}}} = \sum_{i=1}^{k} \rho_i\beta_i^{SLP}(t)\vec{F}^i_{pre}.$$

The analysis described above was performed with the 1950–80 segment of a data set that extends back to 1901. Since the 1901–49 segment is independent of that used to 'train' the model (14.27), it can be used to validate the model. Figure 14.5 shows both the specified and observed winter mean rainfall averaged over all Iberian stations for this period. The overall upward trend and the low-frequency variations in observed precipitation are well reproduced by the indirect method indicating the usefulness of the technique (14.27) as well as the reality of both the trend and the variations in the Iberian winter precipitation.

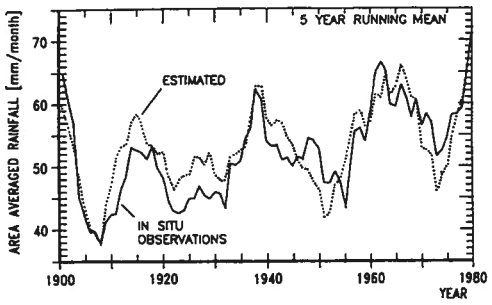


Figure 14.5: *Five-year running mean of winter mean rainfall averaged across Iberian rain gauges. The solid curve is obtained from station data, and the dotted curve is imputed from North Atlantic SLP variations [403].*

14.3.4 North Atlantic SLP and Iberian Rainfall: Downscaling of GCM output. The regression approach described above has an interesting application in climate change studies. GCMs are widely used to assess the impact that increasing concentrations of greenhouse gases might have on the climate system. But, because of their resolution, GCMs do not represent the details of regional climate change well. The *minimum scale* that a GCM is able to resolve is the distance between two neighbouring grid points whereas the *skilful scale* is generally accepted to be four or more grid lengths. The minimum scale in most climate models in the mid 1990s is of the order of 250–500 km so that the skilful scale is at least 1000–2000 km.

Thus the scales at which GCMs produce useful information does not match the scale at which many users, such as hydrologists, require information. *Statistical downscaling* [403] is a possible solution to this dilemma. The idea is to build a statistical model from historical observations that relates large-scale information that can be well simulated by GCMs to the desired regional scale information that can not be simulated. These models are then applied to the large-scale model output.

The following steps must be taken.

1 Identify a regional climate variable $\vec{\mathbf{R}}$ of interest.

2 Find a climate variable $\vec{\mathbf{L}}$ that

- controls $\vec{\mathbf{R}}$ in the sense that there is a statistical relationship between $\vec{\mathbf{R}}$ and $\vec{\mathbf{L}}$ of the form

$$\vec{\mathbf{R}} = \mathcal{G}(\vec{\mathbf{L}}, \vec{\alpha}) + \epsilon \tag{14.28}$$

in which $\mathcal{G}(\vec{\mathbf{L}}, \vec{\alpha})$ represents a substantial fraction of the total variance of $\vec{\mathbf{R}}$. Vector $\vec{\alpha}$ contains parameters that can be used to adjust the fit of (14.28).

- is reliably simulated in a climate model.

3 Use historical realizations $(\vec{\mathbf{r}}_t, \vec{\mathbf{l}}_t)$ of $(\vec{\mathbf{R}}, \vec{\mathbf{L}})$ to estimate $\vec{\alpha}$.

4 Validate the fitted model on independent historical data or by cross-validation (see [18.5.2]).

5 Apply the validated model to GCM simulated realizations of $\vec{\mathbf{L}}$.

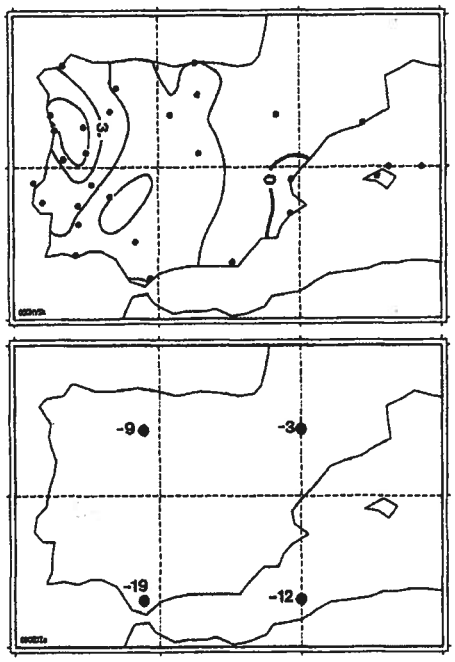


Figure 14.6: *Downscaled and grid point response of Iberian precipitation in a '2×CO_2 experiment' [403].*

This is exactly the process that was followed in the previous subsection. A model (14.27) was constructed that related Iberian rainfall $\vec{\mathbf{R}}$ to North Atlantic SLP $\vec{\mathbf{L}}$ through a simple linear functional. The adjustable parameters $\vec{\alpha}$ consisted of the canonical correlation patterns $\vec{F}^1_{pre}$ and

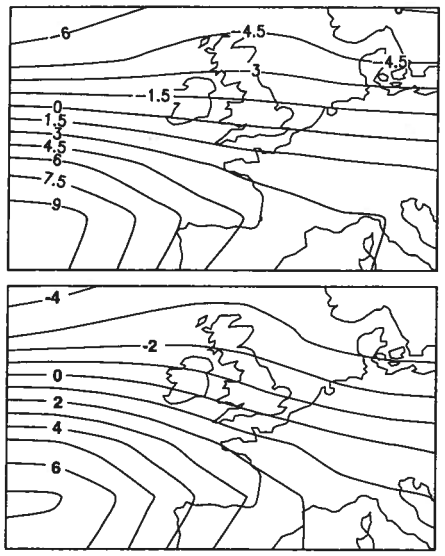


Figure 14.7: *Principal Prediction Patterns* $\vec{F}_0^1$ *(top) and* $\vec{F}_\Delta^1$ *with* $\Delta = 3$ *days (bottom) for the North Atlantic / European Sector daily winter SLP. From Dorn and von Storch [103].*

$\vec{F}_{SLP}^1$ and the canonical correlation ρ_1. These parameters were estimated from 1950 to 1980 data. Observations before 1950 have been used to validate the model.

Downscaling model (14.27) was applied to the output of a '$2\times CO_2$' experiment performed with a GCM. Figure 14.6 compares the 'downscaled' response to doubled CO_2 with the model's grid point response. The latter suggests that there will be a marked decrease in precipitation over most of the Peninsula whereas the downscaled response is weakly positive. The downscaled response is physically more reasonable than the direct response of the model.

14.3.5 Principal Prediction Pattern of North Atlantic / European SLP. Dorn and von Storch [103] used the Principal Prediction Pattern (PPP) analysis technique to study the synoptic predictability of sea-level pressure (SLP) over the eastern North Atlantic and Western Europe. This particular field was used because a rich data set, consisting of daily analysis since approximately 1900, was available for determining the skill of the PPP model.

The PPP analysis was performed with daily winter SLP maps for 1958–88. The dimensionality of the problem was reduced by projecting the maps onto the first eight EOFs of daily winter SLP. Analyses were performed for lags $\tau = 1, \ldots, 5$ days, but we discuss only the $\tau = 3$ days results below.

The first pair of PPPs is shown in Figure 14.7. The patterns are normalized such that the variance of the coefficient of $\vec{F}_0^1$ is 1, and that of the coefficient of $\vec{F}_3^1$ is $1/\sqrt{\rho_1}$. With this normalization, the coefficient for the regression of the $\vec{F}_0^1$-coefficient on the $\vec{F}_3^1$-coefficient is the identity. Also, the patterns are scaled so that if the initial state is a multiple of $\vec{F}_0^1$, then the best predictor is the same multiple of $\vec{F}_3^1$. Patterns $\vec{F}_0^1$ and $\vec{F}_3^1$ are rather similar indicating that the analysis has selected the regional SLP mode that is most persistent on synoptic time scales. The reduction of the magnitude by about 1/3 indicates that this persistence goes with some damping. Thus, the forecast incorporated in this pair of patterns implies constancy in the pattern, but a reduction of the intensity, i.e., 'damped persistence'.

This statement also holds for the other patterns and is further supported by comparing the forecast skill, as given by the anomaly correlation coefficient[11] between the true SLP field and the field predicted by either PPP or persistence (Figure 14.8). The skill of the two forecast schemes is practically identical and exhibits the characteristic decay with increasing lag. Thus, the PPP forecast is no more skilful than the simpler 'competitor' persistence.

However, the PPP forecast scheme should not be dismissed out of hand. By conditioning on the proportion of spatial variance represented by $\vec{F}_0^1$, the PPP forecast was found to be more skilful when the proportion is large (Figure 14.8, bottom). Thus the PPP scheme also gives a forecast of forecast skill.

The utility of the PPP technique needs further exploration and the user is advised to examine all results obtained with this technique critically. In particular, surprisingly good results may be generated by using short time series or by failing to adequately reduce the degrees of freedom of the problem.

14.4 Redundancy Analysis

14.4.1 Introduction. So far, we have identified pairs of patterns by maximizing the correlation

[11] This measure of skill is explained in detail in [18.2.9]. Roughly speaking, it is the mean spatial correlation between the forecast and the verifying field.

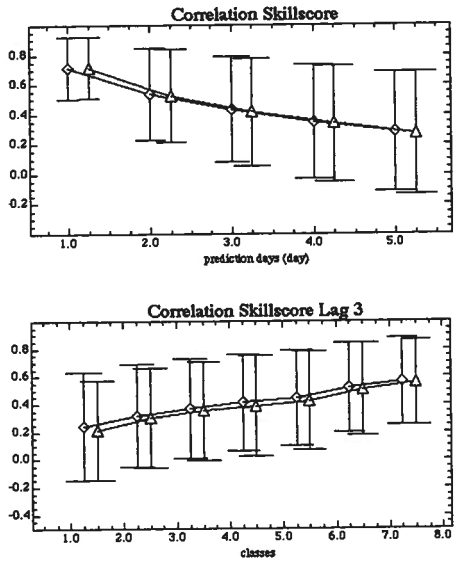


Figure 14.8: *Anomaly correlation coefficient of the PPP forecast (diamonds) and of persistence (triangles).The vertical bars indicate $\pm\sigma$ bands, as estimated from all forecast prepared for the winter days from 1900 until 1990. For better readability, the numbers for the two forecast schemes, persistence and PPP, are shifted horizontally.*
Top: For lags $\tau = 1, \ldots, 5$ days.
Bottom: For lag $\tau = 3$ days. The anomaly correlation coefficients were classified according to the proportion of variance of the initial SLP field described by the PPP (bottom). Class 1 contains cases with proportions in the range [0.0, 0.4], *class 2 contains cases with proportions in* (0.4, 0.5], *and so on up to class 7, which contains cases with proportions in* (0.9, 1].

between the corresponding pattern coefficients. We then demonstrated how regression techniques can be used to specify or forecast the value of the pattern coefficients of one of the fields from those of the other field. This regression problem is generically non-symmetric because the objective is to maximize the variance of the predictand that can be represented. Properties of the predictor patterns, such as the amount of variance they represent, are irrelevant to the regression problem. Hence, there is a mismatch between CCA, which treats variables equally, and regression analysis, which focuses primarily on the predictand.

The 'redundancy analysis' technique directly addresses this problem by identifying patterns that are strongly linked through a regression model. Patterns are selected by maximizing predictand variance. This technique was developed in the late 1970s but apparently has not been introduced in climate research literature.

Here we present the *redundancy analysis* as suggested by Tyler [376]. *Note that very little experience has been collected with this technique in the field of climate research. Therefore, the technique should be applied with great care, and results should be appraised critically.*

14.4.2 Redundancy Index. Let us consider a pair of random vectors $(\vec{\mathbf{X}}, \vec{\mathbf{Y}})$ with dimensions m_X and m_Y. Let us assume further that there is a linear operator represented by a $m_X \times k$ matrix $\mathcal{Q}_k$. How much variance in $\vec{\mathbf{Y}}$ can be accounted for by a regression of $\mathcal{Q}_k^{\mathrm{T}}\vec{\mathbf{X}}$ on $\vec{\mathbf{Y}}$?[12] We assume, without loss of generality, that the expected value of both $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ is zero.

The regression model that relates $\mathcal{Q}_k^{\mathrm{T}}\vec{\mathbf{X}}$ is given by

$$\vec{\mathbf{Y}} = \mathcal{R}(\mathcal{Q}_k^{\mathrm{T}}\vec{\mathbf{X}}) + \vec{\epsilon}, \tag{14.29}$$

where $\mathcal{R}$ is an $m_Y \times k$ matrix of regression coefficients. The variance represented by $(\mathcal{Q}_k^{\mathrm{T}}\vec{\mathbf{X}})$ is maximized when

$$\mathcal{R} = \boldsymbol{\Sigma}_{Y,\mathcal{Q}X}\left(\boldsymbol{\Sigma}_{\mathcal{Q}X,\mathcal{Q}X}\right)^{-1}, \tag{14.30}$$

where

$$\boldsymbol{\Sigma}_{Y,\mathcal{Q}X} = \mathrm{Cov}\left(\vec{\mathbf{Y}}, \mathcal{Q}_k^{\mathrm{T}}\vec{\mathbf{X}}\right) = \boldsymbol{\Sigma}_{YX}\mathcal{Q}_k \tag{14.31}$$

$$\boldsymbol{\Sigma}_{\mathcal{Q}X,\mathcal{Q}X} = \mathcal{Q}_k^{\mathrm{T}}\boldsymbol{\Sigma}_{XX}\mathcal{Q}_k. \tag{14.32}$$

Tyler [376] called the proportion of variance represented by the regression (14.29) the *redundancy index* and labelled it

$$R^2(\vec{\mathbf{Y}} : \mathcal{Q}_k^{\mathrm{T}}\vec{\mathbf{X}}) = \frac{\mathrm{tr}\left(\mathrm{Cov}(\vec{\mathbf{Y}}, \vec{\mathbf{Y}}) - \mathrm{Cov}(\vec{\mathbf{Y}} - \hat{\vec{\mathbf{Y}}}, \vec{\mathbf{Y}} - \hat{\vec{\mathbf{Y}}})\right)}{\mathrm{tr}\left(\mathrm{Cov}(\vec{\mathbf{Y}}, \vec{\mathbf{Y}})\right)} \tag{14.33}$$

where $\hat{\vec{\mathbf{Y}}} = \mathcal{R}(\mathcal{Q}_k^{\mathrm{T}}\vec{\mathbf{X}})$ is the estimated value of $\vec{\mathbf{Y}}$. The motivation of this wording is that it is a measure of how *redundant* the information in $\vec{\mathbf{Y}}$ is if one already has the information provided by $\vec{\mathbf{X}}$.

[12] The number of columns (patterns) in $\mathcal{Q}_k$ is smaller than the dimension of $\vec{\mathbf{X}}$ in most practical situations, so that $k < m_X$ or even $k \ll m_X$. Thus, the operation $\vec{\mathbf{X}} \to \mathcal{Q}_k^{\mathrm{T}}\vec{\mathbf{X}}$ represents a reduction of the phase space of $\vec{\mathbf{X}}$, as in all the other cases we have discussed in this and the previous chapter.

The numerator is the trace (sum of main diagonal elements) of the matrix

$$\begin{aligned}&\Sigma_{YY} - (\Sigma_{YY} + \Sigma_{\hat{Y}\hat{Y}} - 2\Sigma_{Y\hat{Y}})\\ &= -\mathcal{R}\Sigma_{\mathcal{Q}X,\mathcal{Q}X}\mathcal{R}^{\mathrm{T}} + 2\Sigma_{Y,\mathcal{Q}X}\mathcal{R}^{\mathrm{T}}.\end{aligned}$$

Using (14.31)–(14.33), and simplifying, we find that

$$R^2(\vec{\mathbf{Y}} : \mathcal{Q}_k^{\mathrm{T}}\vec{\mathbf{X}}) = \frac{\mathrm{tr}\left(\Sigma_{YX}\mathcal{Q}_k\left(\mathcal{Q}_k^{\mathrm{T}}\Sigma_{XX}\mathcal{Q}_k\right)^{-1}\mathcal{Q}_k^{\mathrm{T}}\Sigma_{XY}\right)}{\mathrm{tr}\left(\Sigma_{YY}\right)}. \tag{14.34}$$

14.4.3 Invariance of the Redundancy Index to Linear Transformations. The redundancy index has a number of interesting properties. One of these is its invariance to orthonormal transformations of $\vec{\mathbf{Y}}$: if $\mathcal{A}$ is orthonormal, then

$$R^2(\mathcal{A}\vec{\mathbf{Y}} : \mathcal{Q}_k^{\mathrm{T}}\vec{\mathbf{X}}) = R^2(\vec{\mathbf{Y}} : \mathcal{Q}_k^{\mathrm{T}}\vec{\mathbf{X}}). \tag{14.35}$$

The significance of this property comes from the fact that we may identify any orthonormal transformation with a linear transformation that conserves variance. Relationship (14.35) does not hold for general non-singular matrices, in particular not for transformations that change the variance since the proportion of captured variance changes when the variance of $\vec{\mathbf{Y}}$ is changed.

On the other hand, any square non-singular matrix $\mathcal{Q}_{m_X}$ used to transform the specifying variable $\vec{\mathbf{X}}$ has also no effect on the redundancy index. In that case, $(\mathcal{Q}_{m_X})^{-1}$ exists and $(\mathcal{Q}_{m_X}^{\mathrm{T}}\Sigma_{XX}\mathcal{Q}_{m_X})^{-1} = \mathcal{Q}_{m_X}^{-1}\Sigma_{XX}^{-1}(\mathcal{Q}_{m_X}^{\mathrm{T}})^{-1}$ in the numerator of (14.34), so that

$$R^2(\vec{\mathbf{Y}} : \mathcal{Q}_{m_X}^{\mathrm{T}}\vec{\mathbf{X}}) = R^2(\vec{\mathbf{Y}} : \vec{\mathbf{X}}). \tag{14.36}$$

The implication of (14.36) is that the coordinate system in which the random vector $\vec{\mathbf{X}}$ is given does not matter, so long as it describes the same linear space. This is a favourable property since the information contained in $\vec{\mathbf{X}}$ about $\vec{\mathbf{Y}}$ should not depend on the specifics of the presentation of $\vec{\mathbf{X}}$, such as the metric used to measure the components of $\vec{\mathbf{X}}$, or the order of its components.

However, if the linear transformation $\mathcal{Q}_k$ maps the m_X-dimensional variable $\vec{\mathbf{X}}$ onto a k-dimensional variable $\vec{\mathbf{X}}'_k = \mathcal{Q}_k^{\mathrm{T}}\vec{\mathbf{X}}$, the new variable contains less information about $\vec{\mathbf{Y}}$, so that

$$\begin{aligned}R^2(\vec{\mathbf{Y}} : \vec{\mathbf{X}}'_k) &\leq R^2(\vec{\mathbf{Y}} : \vec{\mathbf{X}}'_{k+1}) \\ &\leq R^2(\vec{\mathbf{Y}} : \vec{\mathbf{X}}'_{m_X}) = R^2(\vec{\mathbf{Y}} : \vec{\mathbf{X}})\end{aligned} \tag{14.37}$$

provided that 'column spaces'[13] of $\mathcal{Q}_k$, $\mathcal{Q}_{k+1}$, and $\mathcal{Q}_{m_X}$ are nested and $\mathcal{Q}_{m_X}$ is invertible. If, for all k, $\mathcal{Q}_{k+1}$ is constructed by adding a column to $\mathcal{Q}_k$, then inequality (14.37) simply reflects the fact that the regression on $\vec{\mathbf{Y}}$ has k predictors in the case of $\vec{\mathbf{X}}'_k$, and the same k predictors plus one more in the case of $\vec{\mathbf{X}}'_{k+1}$.

For a given transformation $\mathcal{Q}_k$, again only the subspace spanned by the columns of $\mathcal{Q}_k$ matters. That is, for any invertible $k \times k$ matrix $\mathcal{L}$, we find

$$R^2\left(\vec{\mathbf{Y}} : \mathcal{L}^{\mathrm{T}}(\mathcal{Q}_k^{\mathrm{T}}\vec{\mathbf{X}})\right) = R^2(\vec{\mathbf{Y}} : \mathcal{Q}_k^{\mathrm{T}}\vec{\mathbf{X}}). \tag{14.38}$$

Thus, the redundancy index for two variables is a function of the subspace the variable $\vec{\mathbf{X}}$ is projected upon, and the way in which $\vec{\mathbf{Y}}$ is scaled.

Since R^2 does not depend on the specific coordinates of the variable $\vec{\mathbf{X}}$ and $\vec{\mathbf{X}}'_k$, we may assume that the columns of $\mathcal{Q}_k$ are chosen to be orthogonal with respect to $\vec{\mathbf{X}}$; that is,

$$\vec{q}^{\,k\mathrm{T}}\Sigma_{XX}\vec{q}^{\,j} = 0 \tag{14.39}$$

for any $k \neq j$. Then

$$R^2(\vec{\mathbf{Y}} : \mathcal{Q}_k^{\mathrm{T}}\vec{\mathbf{X}}) = \sum_{j=1}^{k} R^2(\vec{\mathbf{Y}} : \vec{q}^{\,j\mathrm{T}}\vec{\mathbf{X}}), \tag{14.40}$$

which may be seen as a special version of (14.37). Note that (14.39) is fulfilled if the vectors $\vec{q}^{\,j}$ are the EOFs of $\vec{\mathbf{X}}$.

14.4.4 Redundancy Analysis. The theory behind *redundancy analysis*, as put forward by Tyler [376], confirms the existence of a non-singular transformation $\mathcal{B} = (\vec{b}^1|\vec{b}^2|\cdots|\vec{b}^{m_X})$ so that the index of redundancy (i.e., the amount of $\vec{\mathbf{Y}}$-variance explained through the regression of $\mathcal{B}_k^{\mathrm{T}}\vec{\mathbf{X}}$ on $\vec{\mathbf{Y}}$) is maximized for any $k = 1, \ldots, \min(m_X, m_Y)$. Matrix $\mathcal{B}_k$ contains the first k columns of $\mathcal{B}$.

Thus redundancy analysis determines the k-dimensional subspace that allows for the most efficient regression on $\vec{\mathbf{Y}}$. Since we are free to choose the coordinates of this subspace, we may use a linear basis with k orthogonal patterns that satisfies (14.39), so that the redundancy index may be expressed specifically as (14.40).

The following theorem identifies a second set of patterns, $\mathcal{A} = (\vec{a}^1|\vec{a}^2|\cdots|\vec{a}^k)$, that represent an orthogonal partitioning of the variance of $\vec{\mathbf{Y}}$ that is accounted for by the regression of $\vec{\mathbf{X}}$ on $\vec{\mathbf{Y}}$. More specifically, the regression maps the subspace

[13] The *column space* of a matrix $\mathcal{Q}$ is the vector space spanned by the columns of $\mathcal{Q}$.

represented by $\vec{\mathbf{X}}'_k$ onto the space spanned by the first k columns of $\mathcal{A}$.

The following subsections describe the mathematics required for the determination of matrices $\mathcal{A}$ and $\mathcal{B}$. The theorems are taken from Tyler's paper [376].

14.4.5 The Redundancy Analysis Transformations. *For any random vectors $\vec{\mathbf{Y}}$ of dimension m_Y and $\vec{\mathbf{X}}$ of dimension m_X, there exists an orthonormal transformation $\mathcal{A}$ and a non-singular transformation $\mathcal{B}$ such that*

$$\mathrm{Cov}(\mathcal{B}^{\mathrm{T}}\vec{\mathbf{X}}, \mathcal{B}^{\mathrm{T}}\vec{\mathbf{X}}) = \mathcal{I} \tag{14.41}$$

$$\mathrm{Cov}(\mathcal{A}^{\mathrm{T}}\vec{\mathbf{Y}}, \mathcal{B}^{\mathrm{T}}\vec{\mathbf{X}}) = \mathcal{D} \tag{14.42}$$

where $\mathcal{D}$ is an $m_Y \times m_X$ matrix with elements $d_{ij} = 0$ for $i \neq j$ and diagonal elements $d_{jj} = \sqrt{\lambda_j}$ for $j \leq \min(m_X, m_Y)$.

The proof, which is detailed in Appendix M, revolves around two eigen-equations:

$$\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\vec{a}^{\,j} = \lambda_j \vec{a}^{\,j} \tag{14.43}$$

$$\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YX}\vec{b}^{\,j} = \lambda_j \vec{b}^{\,j}. \tag{14.44}$$

Both equations have the same positive eigenvalues λ_j, and the eigenvectors $\vec{a}^{\,j}$ and $\vec{b}^{\,j}$ belonging to the same nonzero eigenvalue λ_j are related through

$$\vec{b}^{\,j} = \frac{1}{\sqrt{\lambda_j}}\Sigma_{XX}^{-1}\Sigma_{XY}\vec{a}^{\,j}. \tag{14.45}$$

The matrices $\mathcal{A}$ and $\mathcal{B}$, which are composed of eigenvectors $\vec{a}^{\,j}$ and $\vec{b}^{\,j}$, respectively, are the only matrices that satisfy the requirements of the theorem.

From the computational point of view, it is advisable to solve the eigenproblem with the Hermitian matrix (14.43), then use the identity (14.45). Since (14.43) is a Hermitian problem, all eigenvectors $\vec{a}^{\,j}$ are real valued, and since (14.45) involves only real matrices, the 'patterns' $\vec{b}^{\,j}$ are also real valued.

14.4.6 Theorem: Optimality of the Redundancy Transformation. The significance of the redundancy transformation originates from the following theorem given by Tyler [376]:

The redundancy index $R^2(\vec{\mathbf{Y}} : \mathcal{Q}_k^{\mathrm{T}}\vec{\mathbf{X}})$ is maximized by setting $\mathcal{Q}_k = \mathcal{B}_k$, where $\mathcal{B}_k$ is the $m_x \times k$ matrix that contains the k eigenvectors satisfying (14.42) that correspond to the k largest eigenvalues.

Note that the statement holds for all $k \leq m_X$. Thus, among all possible single patterns $\vec{q}$, the eigenvector $\vec{b}^{\,1}$ belonging to the largest eigenvalue of the matrix $\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YX}$ provides the maximum information, in a linear sense, about the variance of $\vec{\mathbf{Y}}$:

$$R^2(\vec{\mathbf{Y}} : \vec{q}^{\,\mathrm{T}}\vec{\mathbf{X}}) \leq R^2(\vec{\mathbf{Y}} : \vec{b}^{\,1\mathrm{T}}\vec{\mathbf{X}}) \tag{14.46}$$

for any m_X-dimensional vector $\vec{q}$. Moreover, by equations (14.41) and (14.40), the index of redundancy takes a particularly simple form,

$$R^2(\vec{\mathbf{Y}} : \mathcal{B}_k^{\mathrm{T}}) = \sum_{j=1}^{k} R^2(\vec{\mathbf{Y}} : \vec{b}^{\,j\mathrm{T}}\vec{\mathbf{X}}). \tag{14.47}$$

Also, note that inequality (14.46) may be generalized to

$$\sum_{j=1}^{k} R^2(\vec{\mathbf{Y}} : \vec{q}^{\,j\mathrm{T}}\vec{\mathbf{X}}) \leq \sum_{j=1}^{k} R^2(\vec{\mathbf{Y}} : \vec{b}^{\,j\mathrm{T}}\vec{\mathbf{X}}) \tag{14.48}$$

for any set of vectors $\vec{q}^{\,1}, \ldots, \vec{q}^{\,k}$.

14.4.7 The Role of Matrix $\mathcal{A}$. Since $\mathcal{B} = (\vec{b}^{\,1}|\cdots|\vec{b}^{\,m_X})$ is non-singular, random vector $\vec{\mathbf{X}}$ can be expanded in the usual manner as

$$\vec{\mathbf{X}} = \sum_{j=1}^{m_X} (\vec{\mathbf{X}}^{\mathrm{T}}\vec{b}^{\,j})\vec{p}^{\,j}, \tag{14.49}$$

where the *adjoint patterns* $\mathcal{P} = (\vec{p}^{\,1}|\cdots|\vec{p}^{\,m_X})$ are given by $\mathcal{P}^{\mathrm{T}} = \mathcal{B}^{-1}$. When re-expressed in matrix-vector form, equation (14.49) simply reads as

$$\vec{\mathbf{X}} = \mathcal{P}\mathcal{B}^{\mathrm{T}}\vec{\mathbf{X}}.$$

Similarly, since $\mathcal{A}$ is orthonormal, the part of $\vec{\mathbf{Y}}$ that can be represented by $\vec{\mathbf{X}}$, that is, $\hat{\vec{\mathbf{Y}}}$, can be expanded as

$$\hat{\vec{\mathbf{Y}}} = \mathcal{A}\mathcal{A}^{\mathrm{T}}\hat{\vec{\mathbf{Y}}} = \sum_j (\hat{\vec{\mathbf{Y}}}^{\mathrm{T}}\vec{a}^{\,j})\vec{a}^{\,j}. \tag{14.50}$$

When we regress $\vec{\mathbf{Y}}$ on $\vec{\mathbf{X}}$, we find that $\hat{\vec{\mathbf{Y}}} = \Sigma_{YX}\Sigma_{XX}^{-1}\vec{\mathbf{X}}$. Thus the expansion coefficients in (14.50), the elements of $\mathcal{A}^{\mathrm{T}}\hat{\vec{\mathbf{Y}}}$, are given by

$$\mathcal{A}^{\mathrm{T}}\hat{\vec{\mathbf{Y}}} = \mathcal{A}^{\mathrm{T}}\Sigma_{YX}\Sigma_{XX}^{-1}\vec{\mathbf{X}}.$$

Now, from equations (14.41) and (14.42) we have that $\Sigma_{XX}^{-1} = \mathcal{B}\mathcal{B}^{\mathrm{T}}$ and $\mathcal{A}^{\mathrm{T}}\Sigma_{YX}\mathcal{B} = \mathcal{D}$. Thus

$$\mathcal{A}^{\mathrm{T}}\hat{\vec{\mathbf{Y}}} = \mathcal{A}^{\mathrm{T}}\Sigma_{YX}\mathcal{B}\mathcal{B}^{\mathrm{T}}\vec{\mathbf{X}} = \mathcal{D}\mathcal{B}^{\mathrm{T}}\vec{\mathbf{X}}.$$

Hence the expansion coefficients in (14.50) are given by

$$\hat{\vec{Y}}^{\mathrm{T}}\vec{a}^{\,j} = \sqrt{\lambda_j}\,\vec{X}^{\mathrm{T}}\vec{b}^{\,j}. \tag{14.51}$$

Considering both (14.49) and (14.50), we see that the regression maps variations in the amplitude of $\vec{X}$ patterns $\vec{p}^{\,j}$ onto variations in the amplitude of $\hat{\vec{Y}}$ patterns $\vec{a}^{\,j}$. On average, $\vec{Y}_j = \sqrt{\lambda_j}\,\vec{a}^{\,j}$ when $\vec{X}_j = \vec{p}^{\,j}$ (cf. (14.49) and (14.51)). It is easily shown that the patterns themselves are related by[14]

$$\mathcal{A}\mathcal{D} = \Sigma_{YX}\Sigma_{XX}^{-1}\mathcal{P}.$$

That is, the $\vec{X}$-patterns are transformed into scaled versions of the $\hat{\vec{Y}}$ patterns by the regression operator.

Thus, redundancy analysis offers a number of useful insights. First, it helps us to identify an efficient way of specifying a maximum of variance in one random vector from the information provided by another vector. It also guides us in finding those components of the specifying variable that contain the most information about the variable to be specified. Finally, it offers pairs of patterns that are mapped onto each other. If we observe the pattern $\vec{p}^{\,j}$ in the specifying field, then the likelihood of observing pattern $\vec{a}^{\,j}$ in the field to be specified is increased.

If we consider the full $\vec{X}$-space, we find that

$$\Sigma_{\hat{Y}\hat{Y}} = \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}. \tag{14.52}$$

When comparing this expression with the eigenproblem (14.43), it becomes obvious that the $\vec{a}$-vectors are the EOFs of $\hat{\vec{Y}}$. Thus the $\vec{a}^{\,1}$ coefficient accounts for the largest amount of $\vec{Y}$ variance (i.e., λ_1), $\vec{a}^{\,2}$ accounts for the second largest amount of variance λ_2, and so on. The total variance of the regressed vector $\hat{\vec{Y}}$ is $\sum_j \lambda_j$. Since $\Sigma_{Y\hat{Y}} = \Sigma_{\hat{Y}\hat{Y}}$, we have

$$\begin{aligned} R^2(\vec{Y} : \hat{\vec{Y}}) &= R^2(\vec{Y} : \vec{X}) \\ &= \frac{\mathrm{tr}(\Sigma_{\hat{Y}\hat{Y}})}{\mathrm{tr}(\Sigma_{YY})} = \frac{\sum_j \lambda_j}{\mathrm{tr}(\Sigma_{YY})}. \end{aligned} \tag{14.53}$$

When we truncate (14.49) to the k components of $\vec{X}$ that carry the most information about $\vec{Y}$, we find that

$$R^2(\vec{Y} : \hat{\vec{Y}}) = R^2(\vec{Y} : \mathcal{B}_k^{\mathrm{T}}\vec{X}) = \frac{\sum_{j=1}^{k} \lambda_j}{\mathrm{tr}(\Sigma_{YY})}.$$

[14] The proof is straightforward:
$\mathcal{R}\mathcal{P} = \Sigma_{YX}\Sigma_{XX}^{-1}(\mathcal{B}^{\mathrm{T}})^{-1} = \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XX}\mathcal{B} = \Sigma_{YX}\mathcal{B} = \mathcal{A}\mathcal{D}$.

14.4.8 Comparison with CCA. Let us now consider the special case in which Σ_{XX} and Σ_{YY} are both identity matrices. Then $\mathcal{B}$ and $\mathcal{P}$ are also identity matrices, and the regressed patterns $\vec{a}$, the EOFs $\hat{\vec{Y}}$, are the eigenvectors of $\Sigma_{YX}\Sigma_{XY}$. That is, $\vec{X}$ provides the most information about the component of $\vec{Y}$ that lies in the $\vec{a}^{\,1}$ direction, where $\vec{a}^{\,1}$ is the first eigenvector of $\Sigma_{YX}\Sigma_{XY}$. The best predictor of this component is $\vec{X}^{\mathrm{T}}\Sigma_{XY}\vec{a}^{\,1}$.

When we perform CCA on the same system we must solve the paired eigenvalue problem

$$\begin{aligned} \Sigma_{XY}\Sigma_{YX}\vec{f}_X &= \lambda\vec{f}_X \\ \Sigma_{YX}\Sigma_{XY}\vec{f}_Y &= \lambda\vec{f}_Y. \end{aligned}$$

The first pair of eigenvectors of this system is given by $\vec{f}_X = \Sigma_{XY}\vec{a}^{\,1}$, and $\vec{f}_Y = \vec{a}^{\,1}$, indicating that $\vec{X}^{\mathrm{T}}\Sigma_{XY}\vec{a}^{\,1}$ is the $\vec{X}$-component most strongly correlated with $\vec{Y}^{\mathrm{T}}\vec{a}^{\,1}$.

Thus redundancy analysis and CCA are equivalent in this special case: both identify the same $\vec{X}$ and $\vec{Y}$ directions.

In general, however, the methods are not equivalent. Redundancy analysis finds the best predicted (or specified) components of $\vec{Y}$ by finding the eigenvectors $\vec{a}$ of

$$\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$$

and then finding the patterns $\vec{p}$ of $\vec{X}$-variations that carry this information. CCA, on the other hand, finds the most strongly correlated components of $\vec{Y}$ by finding the eigenvectors $\Sigma_{YY}^{-1/2}\vec{f}_Y$ of

$$(\Sigma_{YY}^{-1/2})^{\mathrm{T}}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1/2}.$$

That is, CCA does redundancy analysis on $\vec{Y}' = (\Sigma_{YY}^{-1/2})^{\mathrm{T}}\vec{Y}$, the random vector that is obtained by projecting $\vec{Y}$ onto its EOFs and scaling each component by its standard deviation. We can therefore anticipate that the two techniques will produce similar results if $\vec{Y}$ is projected onto a small number of EOFs with similar eigenvalues.

14.4.9 Example: Interdecadal Variability of Intramonthly Percentiles of Significant 'Brent' Wave Height. We now describe an application in which we use redundancy analysis to specify monthly wave height statistics at the Brent oil field, located northeast of Scotland in the North Atlantic at (61° N, 1.5° E). Wave height (sea state) data are available from visual assessments made on ships of opportunity, at light houses, from wave rider buoys, and shipborne instruments at ocean weather stations. Also, wave height maps

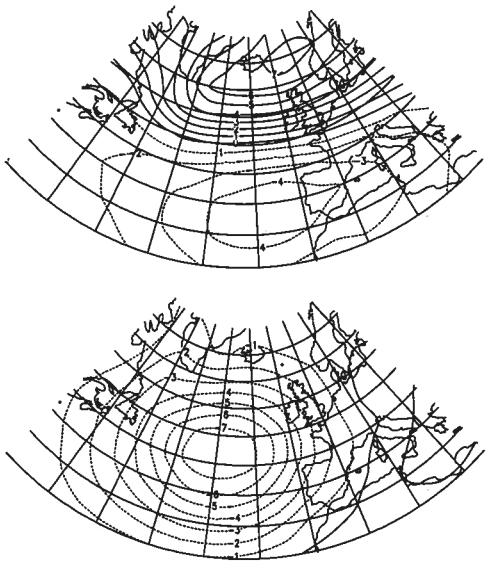


Figure 14.9: *First two monthly mean air-pressure anomaly distributions $\widehat{\vec{p}}^k$ identified in a redundancy analysis as being most strongly linked to simultaneous variations of the intramonthly percentiles of significant wave height in the Brent oil field (61° N, 1.5° E; northeast of Scotland).*

have been constructed from wind analyses for the purpose of ship routing (Bouws et al. [57]). These data are sparse and suffer from various inhomogeneities. Also, the records are generally too short to allow an assessment of changes during the past century.

Thus, observational data alone do not contain sufficient information about the interdecadal variability of wave statistics. One solution is a combined statistical/dynamical reconstruction of the past that uses a dynamical wave model. The model is forced with recent wind data that are believed to be fairly reliable and not strongly affected by improving analysis techniques.[15] The wave heights derived from the hindcast simulation are treated as observations and are used to build a statistical model linking the wave heights to surface air pressure. Finally, the resulting

[15]Note that the homogeneity of weather maps and their surface winds is difficult to assess. Analysis system improvements can introduce artificial signals, such as increasing frequencies of extreme events, into the hindcast. Improved analyses procedures, be it more or better observations or more intelligently designed dynamical and statistical analysis tools, lead to the emergence of more details in weather maps and, therefore, larger extremes.

	Wave height percentile		
	50%	80%	90%
$\widehat{\vec{a}}^1$	−81	−107	−114
$\widehat{\vec{a}}^2$	32	2	−25

Table 14.2: *The vectors $\widehat{\vec{a}}^k$ of anomalous intramonthly percentiles of significant wave height are given as rows in the following table. Units: cm.*

statistical model is fed with the observed air pressure from the beginning of the century onward, thereby producing a plausible estimate of wave height statistics for the entire century. The statistical model is presented below.

In this case we bring together 'apples' and 'oranges', that is, two vector quantities that are not directly linked. One vector time series, $\vec{\mathbf{X}}_t$, represents the winter (DJF) monthly mean surface air-pressure distributions in the North Atlantic. The other vector time series, $\vec{\mathbf{Y}}_t$, is a three-dimensional random vector consisting of the 50th, 80th, and 90th percentiles of the *intramonthly distributions* of significant wave height[16] in the *Brent* oil field at (61° N, 1.5° E). Both vector time series are assumed to be centred, so that the air-pressure values and percentiles are deviations from their respective long-term means.

The monthly mean of North Atlantic SLP is indirectly linked to the intramonthly percentiles, since storms affect both the monthly mean air-pressure distribution and the distribution of wave heights within a month at a specific location. Of course, the storm activity may also be seen as being conditioned by the monthly mean state.

The daily wave height data are taken from a 40-year 'hindcast' simulation (Günther et al. [153]). The following analysis assumes that the hindcasts and windfield analyses both represent the real world well enough for statistical relationships between the wave and wind fields on the monthly time scale to be reliably diagnosed.

A redundancy analysis of the two vector time series is performed to detect the dominant coupled anomaly patterns in the mean air pressure and in the intramonthly wave height percentiles. The SLP patterns $\widehat{\vec{p}}^1$ and $\widehat{\vec{p}}^2$ are shown in Figure 14.9 and the corresponding intramonthly percentiles $\widehat{\vec{a}}^1$ and $\widehat{\vec{a}}^2$ are listed in Table 14.2. The time coefficients are normalized to unit variance so that

[16]Significant wave height is a physical parameter that describes the wave field on the sea surface. The word 'significant' does not imply a significance test in this context. See [3.2.4].

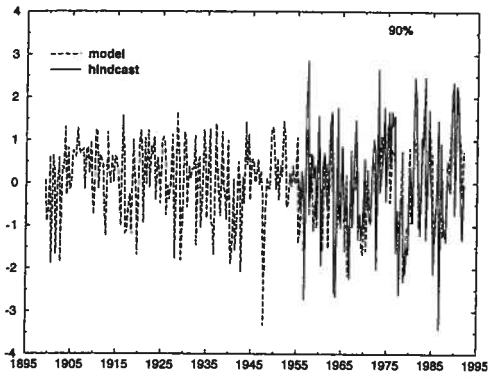


Figure 14.10: *Reconstructed (dashed line) and hindcasted (continuous line; 1955–94) anomalies of the 90th percentile of significant wave height at in the Brent oil field. Units: m.*

the three components of $\widehat{\vec{a}}^k$ may be interpreted as typical anomalies that occur when the pressure field anomalies are given by $\frac{1}{\sqrt{\lambda_k}}\widehat{\vec{p}}^k$.

The pattern $\vec{a}^1$ accounts for 94% of the variance of $\vec{\mathbf{Y}}$, and $\vec{a}^2$ for 5%. The correlation between the coefficient time series of the first pair of vectors is 0.84 while that between the second pair is only 0.08. Thus, the first pair establishes a regression representing $R^2(\vec{\mathbf{Y}} : \mathcal{B}_1^T\vec{\mathbf{X}}) = 94\% \times 0.84^2 = 66\%$ of the variance of $\vec{\mathbf{Y}}$, whereas the second pair represents only $5\% \times 0.08^2 < 0.1\%$ of variance. Thus the redundancy index for $k = 1$ (0.66) can not be usefully increased by adding a second vecor.

The first air-pressure pattern is closely related to the *North Atlantic Oscillation* (see [13.5.5] and Figure 13.6). A weakening of the NAO is associated with a decrease in all three intramonthly percentiles of significant wave height. In effect, this pattern describes a shift of the intramonthly distribution towards smaller waves.

The second pattern describes a mean south-easterly flow across the northern North Sea. The 50th percentile of the significant wave heights is increased by 32 cm, while the 90th percentile is reduced by 25 cm. Thus there is a tendency for the wave height distribution to be widened when pressure anomaly pattern $\widehat{\vec{p}}^2$ prevails. The reversed pattern goes with a narrowed intramonthly distribution of wave heights. This pair of patterns accounts for only 5% of the predictable wave height variance.

The regression model incorporated in the redundancy analysis was used to estimate the time series of the percentiles of significant wave height in the Brent oil field from the observed monthly mean air pressure anomaly fields between 1899 and 1994. The last 40 years may be compared with the hindcast data, whereas the first 50 years represent our best guess and can not be verified at this time. The 90th percentiles of the reconstructed wave height time series for 1899–94 and the corresponding hindcasted time series for 1955–94 are shown in Figure 14.10. The link appears to be strong, as is demonstrated by the correlations and the proportion of described variance, during the overlapping period:

	Wave height percentile		
	50%	80%	90%
Correlation	0.83	0.82	0.77
Described variance	0.70	0.66	0.60

The amount of percentile variance represented by the SLP patterns is consistent with the redundancy index (14.53), which has value 0.66. As with all regression models, the variance of the estimator is smaller than the variance of the original variable. This makes sense, since the details of the wave action in a month are not completely determined by the monthly mean air-pressure field. It is also affected by variations in surface wind that occur on shorter time scales.