

## Chapter 3

# Moving patterns in space-time data sets

Many geophysical observations are in the form of space-time data sets, i.e. time series of a physical variable measured at several spatial locations. They may contain moving patterns / traveling waves of different spatial scales and temporal frequencies. Methods used to study the characteristics of traveling waves in geophysical data are:

Analysis of Fourier coefficients of time series

Cross spectral analysis

Lagged cross correlation analysis

Frequency Domain PCA (FDPCA)

Time Domain Complex PCA (TDCPCA)

### 3.1 Analysis of Fourier coefficients of time series

Consider a space-time function of a physical variable with or without propagating signals embedded in it, observed at discrete points in space and time, denoted by a matrix  $X$  of dimension  $(N, p)$ :

$$\{X(t, i); t = n\Delta t; n = 1, 2, \dots, N; i = 1, 2, \dots, p\} = \{X_{ni}\}$$

where  $N$  denotes number of time steps,  $\Delta t$  is the sampling interval, and  $p$  the number of spatial locations where observations are made. For example, the physical variable could be the SST measured at  $p$  stations over  $N$  time steps. To study the characteristics of moving patterns in such data fields, the given time series at each observation station is expressed in terms of its Fourier coefficients  $a_{mi}$ ,  $b_{mi}$  as [Bendat and Piersol, 1971],

$$X(t, i) = a_{oi}/2 + \sum_{m=1}^{N/2} [a_{mi} \cos(\omega_m t) + b_{mi} \sin(\omega_m t)] \quad (3.1)$$

$$= \sum_{m=1}^{N/2} A_{mi} \cos(\omega_m t - \phi_{mi}) \quad (3.2)$$

where the angular frequency  $\omega_m$  is given by  $\omega_m = 2\pi m/(N\Delta t) = 2\pi f_m$ ,  $f_m$  being the frequency, with its inverse being the period,  $T$ . The amplitude is given by  $A_{mi} = \sqrt{a_{mi}^2 + b_{mi}^2}$  and the phase  $\phi_{mi}$  is given by  $\phi_{mi} = \arctan[b_{mi}/a_{mi}]$  for  $m = 1, 2, \dots, N/2$  and  $i = 1, 2, \dots, p$  and  $a_{oi} = 0$ , when  $X$  are expressed as deviations from their column means.

For any desired angular frequency  $\omega_m$ , we can get information on travelling waves by plotting  $\phi_{mi}$  at all spatial locations and drawing contours of  $\phi_{mi}$  [Preisendorfer, 1988]. The normal to the contours of  $\phi_{mi}$  gives the propagation direction of these waves.

The Fourier coefficients  $a_{mi}$ , and  $b_{mi}$  are the real and imaginary parts of  $(2/N)\hat{X}(\omega_m, i)$ , where  $\hat{X}(\omega_m, i)$ , the Fourier transform the time series  $X(t, i)$  at location  $i$  and angular frequency  $\omega_m$  is given by

$$\hat{X}(\omega_m, i) = \sum_{n=0}^{N-1} X_{ni} \exp\left(\frac{-j2\pi n m}{N}\right) \quad (3.3)$$

where  $j = \sqrt{-1}$

## 3.2 Cross spectral analysis

Consider the data matrix  $X$  of dimension  $(N, p)$ , as in Chapter 3.1, where  $N$  is the number of equally spaced time steps and  $p$  is the number of spatial locations, which may or may not be uniformly gridded.

In the cross-spectral analysis, we select a base series from the data matrix  $X$  (i.e. a column vector of  $X$ ), and compute the cross spectrum between the base series and other time series in the data set. The cross-spectrum is a complex function of frequency  $\omega$  and has the form,

$$S_{bi}(\omega) = C_{bi}(\omega) + j Q_{bi}(\omega) = |S_{bi}(\omega)| \exp[-j\phi_{bi}(\omega)]$$

where  $C_{bi}$  and  $Q_{bi}$  are the co- and quadrature spectra between the base and the  $i$ -th series, and  $j = \sqrt{-1}$ . When applying the cross-spectral method to study the moving patterns in the data set, we use the coherence function  $\gamma_{bi}^2$  defined by,

$$\gamma_{bi}^2(\omega) = |S_{bi}(\omega)|^2 / [S_{bb}(\omega)S_{ii}(\omega)]$$

and the phase function  $\phi_{bi}$  given by,

$$\phi_{bi} = \arctan(Q_{bi}/C_{bi})$$

where  $S_{bi}(\omega)$  is the amplitude of the cross-spectral estimate. By analyzing the coherence and the phase functions, we can gain information on moving patterns in the given data set.

The cross-spectrum  $S_{bi}(\omega)$  is defined in terms of the Fourier Transforms (3.3) of the *base* and the *i*-th series as follows, where  $\langle . \rangle$  indicates averaging over ensembles or frequencies or both:

$$S_{bi}(\omega_m) = \langle \hat{X}^*(\omega_m, b) \hat{X}(\omega_m, i) \rangle \quad (3.4)$$

(3.4) implies that  $S_{ib}(\omega_m) = S_{bi}^*(\omega_m)$ , where  $*$  denotes the complex conjugate. This property of cross-spectrum analysis is used in Chapter 3.4 in forming a  $(p, p)$  Hermitian cross-spectral matrix at each frequency  $\omega_m$ .

Cross-spectral analysis has been used in number of investigations to study moving patterns in geophysical data fields. For example Wallace [1971] used cross spectrum analysis to study tropospheric wave disturbances in the tropical western Pacific. Fernandes et al. [2002], Fernandes [1999] and Fernandes et al. [2000a] have used cross spectrum analysis to determine the wavelength and wave direction of ocean waves ( $1 < T < 30s$ ) and surf beat ( $30s < T < 5minutes$ ) as a function of frequency from array measurements at Duck, North Carolina, U.S.A.

The shortcoming of this technique is the subjectivity in the choice of the base series in the analysis. And the use of a base series produces a bias in favor of the base series, which can produce considerable distortion in the phase and amplitude distribution of the cross spectrum when the coherence is not high.

### 3.3 Lagged cross correlation analysis

Weare and Nasstrom [1982] determined the eigenvectors of lagged covariance matrices to detect the propagating features in the data sets. This approach is expensive and can only resolve propagating phenomena over a limited time span.

### 3.4 Frequency Domain Complex PCA (FDPCA)

This analysis technique was developed in the meteorological context by Wallace and Dickinson [1972]. They discuss the theoretical basis of this technique with reference to a band filtered time series observed at several spatial locations. This analysis technique was subsequently used by Wallace [1972], Wang and Mooers [1977], Michaelson [1982] and others, and it involves computing complex eigenvectors of cross-spectral matrix of a multivariate time series, rather than those of the covariance matrix.

In the present discussion, we will consider the unfiltered data matrix  $X$  of dimension  $(N, p)$ , as in Chapter 3.1, where  $N$  is the number of equally spaced time steps, and  $p$  is the number of spatial locations, which may or may not be uniformly gridded, so that each column of  $X$  is a time series.

At each frequency  $\omega$  form a cross-spectral matrix  $S$  with complex valued elements  $S_{bi}(\omega)$  defined using the Fourier Transforms  $\hat{X}$  of  $X$  by (3.4) for  $b, i = 1, 2, 3, \dots, p$ .  $S$  is a  $(p, p)$  Hermitian matrix and has  $r$  ( $r \leq p$ ) real, non-negative eigen values  $\lambda_k$  and associated complex orthonormal eigen vectors  $V_k$ . These  $p$ -dimensional eigenvectors define the complex spatial modes. The procedure for determining the eigenvalues and eigenvectors of a Hermitian matrix is discussed in the next section.

Michaelson [1982] has described the procedure for determining the principal components  $\xi_k(t)$ , corresponding to the complex spatial mode  $V_k$ . The elements of  $\xi_k$  are complex and satisfy the orthogonality property.

For every eigen mode in this frequency domain analysis, four measures can be defined for quantifying different propagating features in the data field. The amplitude function of the spatial mode  $V_k$  represents the spatial distribution of variability associated with that mode, and the spatial phase function  $\phi_k$  shows the relative phase of fluctuation at various spatial locations. The spatial gradient of  $\phi_k$  provides the local wave numbers. Similarly, the temporal amplitude function  $\xi_k(t)$  and the temporal phase function  $\theta_k(t)$  provide measures of temporal variability associated with each eigen-mode. The time derivative of  $\theta_k(t)$  is a measure of frequency.

### 3.5 Time Domain Complex PCA (TDCPCA)

To describe this analysis, we will consider the same data matrix  $X$  of dimension  $(N, p)$ , as in Chapter 3.1

$$\{X(t, i); t = n\Delta t; n = 1, 2, \dots, N; i = 1, 2, \dots, p\} = \{X_{ni}\}$$

In this method the real data matrix  $X$  to be analyzed is expressed in the complex function notation, which helps in detecting the propagating features within it. To express this discrete data field  $X$  in complex function notation, an imaginary part is added to each element of the matrix  $X$ . The resulting complex matrix is denoted by  $X_c$ , whose imaginary part is formed by the Hilbert transform of  $X$ .

The Hilbert transform of a data field is obtained from its Fourier series expansion. with its phase advanced at every spectral frequency by  $\pi/2$ . More specifically, let the Fourier representation of a time series  $d(t)$  at any spatial location be,

$$d(t) = \sum_{\omega} [a_{\omega} \cos(\omega t) + b_{\omega} \sin(\omega t)]$$

where  $(a_\omega, b_\omega)$  are the Fourier coefficients defined at positive frequencies in the usual manner. Then the Hilbert transform  $d_H(t)$  can be defined as,

$$d_H(t) = \sum_{\omega} [b_{\omega} \cos(\omega t) - a_{\omega} \sin(\omega t)]$$

and the real space-time data field  $d(t)$  in the complex function notation can be written as,

$$D(t) = d(t) + jd_H(t)$$

It can easily be derived from the above equations that

$$D(t) = \sum_{\omega} A_{\omega} \exp(-j\omega t) \quad (3.5)$$

where  $A_{\omega} = a_{\omega} + jb_{\omega}$ . (3.5) is the standard representation of propagating waves in complex function notation. Thus the augmentation of the data matrix with its Hilbert transform enables recognition of propagating waves.

Determination of Hilbert transform is subject to the problems that normally follow Fourier analysis, e.g. aliasing, end effects, etc. To reduce the end effects in the estimates of Hilbert transforms, nearly 10% at both ends of the time series are tapered.

The next step in this analysis is the representation of the complex data matrix  $X_c$  as a product of orthogonal and orthonormal complex matrices. These alternative representations of the data matrices are similar to the PCA and EOF analyses of real matrices, and also the purpose of these analyses are the same, which have been already discussed in Chapter 2.

As in the case of real matrices, the complex data matrix  $X_c$  is expressed as a product in the form,

$$X_c = \xi V^{*T}$$

with  $\xi = U\Gamma$ , or in the form,

$$X_c = U\zeta^{*T}$$

with  $\zeta = V\Gamma$ . In the above decomposition of  $X_c$ , the column vectors of  $\xi$  and  $\zeta$  are referred to as the Complex Principal Components, and they are orthogonal to their respective complex conjugates. Superscripts \* denotes the complex conjugate and  $T$  denotes the transpose of a matrix.  $\Gamma$  is a diagonal matrix of size  $(r, r)$ , whose diagonal elements are the eigenvalues of the cross-correlation matrix of  $X_c$ , and  $r$  is the rank of the matrix  $X_c$ .

$U$  and  $V$ , referred to as Complex Empirical Orthogonal Functions (Complex EOF), are the eigenvectors of the cross-correlation matrices  $[X_c X_c^{*T}]$  and  $[X_c^{*T} X_c]$ . They are determined by SVD

analysis as in the case of real matrices. That is, the complex data matrix  $X_c$  is expressed in the form of a product of three matrices,

$$X_c = U\Gamma V^{*T} \quad (3.6)$$

where  $U$  and  $V$  are orthonormal to their respective complex conjugates and  $\Gamma$  is a diagonal matrix of size  $(r, r)$ .

The advantages and shortcomings of PC and EOF analyses of real data matrices already discussed in Chapter 2 are true in the case of complex-PC and complex-EOF analyses. An additional attribute of the above complex-PC and complex-EOF analyses is that they help in studying the propagating features in space-time data sets of geophysical fields. The physical significance of the complex spatial and temporal empirical modes defined in terms of their amplitude and phase functions has been already discussed in Chapter 3.4.

To determine  $V$  with desired properties, the complex cross correlation matrix  $W$  of  $X_c$  is constructed as,

$$W = X_c^{*T} X_c$$

By construction, the matrix  $W$  is Hermitian, positive definite and of size  $(p, p)$ . Hence  $W$  possesses real positive eigen-values. For computing the eigenvalues and the eigenvectors of the complex matrix  $W$ , it is converted to a real symmetric matrix. If  $W = A + jB$ , then the eigenvalue problem of the complex matrix  $W$  of size  $(pxp)$  defined as,

$$(A + jB)(\alpha + j\beta) = \lambda(\alpha + j\beta)$$

is equivalent to the eigenvalue problem of a real matrix of size  $(2p, 2p)$  defined below:

$$\begin{bmatrix} A & -B \\ B & A \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \lambda \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

where the eigenvector  $V \equiv (\alpha + j\beta)$ . Note that, since  $W$  is Hermitian and  $A^T = A$  and  $B^T = -B$ , the above  $(2p, 2p)$  matrix is real and symmetric. This eigenvalue problem can be solved using the standard methods. Note that if the eigenvalues of  $W$  are  $\lambda_1, \lambda_2, \dots, \lambda_p$  then the  $2p$  eigenvalues of the augmented problem are  $\lambda_1, \lambda_1, \lambda_2, \lambda_2, \dots, \lambda_p, \lambda_p$ , i.e., the eigenvalues are repeated once. The corresponding eigenvectors of the matrix  $W$  are pairs of the form  $(\alpha + j\beta)$  and  $j(\alpha + j\beta)$ . Hence we choose one from each pair of eigenvalue and eigenvector as the desired solution.

Then using the (3.6) the eigenvector  $U$  of the cross-product matrix  $[X_c X_c^{*T}]$  can be determined as  $U = X_c V \Gamma^{-1}$ .

## Chapter 4

# Relationships between data sets

In large-scale ocean-atmospheric experiments we have no control over the parameters, which describe the phenomenon selected for investigation. We cannot, in general isolate and keep one set of parameters constant and vary others to study the effects of their variation on the chosen phenomenon. In such situations, when the measured data sets are large, certain multivariate analyses serve as tools to detect inter-relationships between the observed parameters. But note that any analysis to detect the relationship between the fields will give meaningful results only if there is strong physical reasoning for the existence of dynamic relationships between the fields.

A simple method for examining the coupled variability of two time dependent fields involves determination of correlations between selected time series from data sets. But this involves subjective decisions on the choice of time series for analysis. Some of the objective methods available for the analysis of such data sets are:

Combined Principal Component Analysis (CPCA)

Single-field-based PC/EOF Analysis (SFPCA)

SVD of Covariance matrix between two fields

Canonical Correlation Analysis (CCA)

Barnett and Preisendorfer method

An important study by Bretherton et al. [1992] compares some of the above methods for detecting coupled pattern between two synthetic time dependent fields.

## 4.1 Combined Principal Component Analysis

In this method [Kutzbach, 1967] two or more field variables are combined into a single matrix and Principal Component Analysis is performed on the resulting matrix to identify the relationships between data fields.

For example we will consider two fields  $X$  and  $Y$ . Suppose one set of variables  $X$  is measured at  $p$  stations, the other set  $Y$  at  $q$  stations, and both observed at  $N$  time levels. Then  $X$  is a matrix of size  $(N, p)$ , and  $Y$  is of size  $(N, q)$ . We shall assume that  $p \leq q$ , and that both  $X$  and  $Y$  are expressed as deviations from a suitably defined means; that is, the mean of each time series in  $X$  and  $Y$  has been removed. Each column of  $X$  and  $Y$  contains a time series of the variables at a particular station, and each row contains a map of the measured variable at a given time.

Now we form the combined data matrix  $[X, Y]$  of size  $(N, p + q)$ , viz.,

$$A = [X, Y] = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} & y_{11} & y_{12} & \cdots & y_{1q} \\ x_{21} & x_{22} & \cdots & x_{2p} & y_{21} & y_{22} & \cdots & y_{2q} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Np} & y_{N1} & y_{N2} & \cdots & y_{Nq} \end{bmatrix}$$

Using the PCA of the combined data matrix  $A$ , we may write,

$$A = \xi V_{xy}^T$$

where the columns of  $\xi$  are the orthogonal time functions, and the columns of  $V_{xy}$  are the space functions of the combined field, which are orthonormal vectors of dimension  $(p + q)$ .

An appropriate way to interpret the results of this analysis in terms of correlated signals in the two fields is to examine the spatial patterns of the data fields. The first  $p$  elements of a column of  $V_{xy}$  describe the spatial pattern/mode of variation of the field  $X$ , and the remaining  $q$  elements describe the spatial pattern of the field  $Y$ . That is through PCA of the combined field, the spatial patterns that simultaneously vary in the two fields are identified. The column vectors of  $\xi$  define the time series that describe how each spatial mode oscillates in time. The first spatial mode of  $X$  when multiplied by the first time function gives an estimate  $\hat{X}$  of the field  $X$  that is optimal in the sense that the variance of  $X$  explained by  $\hat{X}$  is maximum. Also the first spatial mode of  $Y$  when multiplied by the first time function gives the optimal estimate  $\hat{Y}$ . Similarly the higher modes along with their time functions explain in the optimal way the variance of the data field left unexplained by the preceding pairs of spatial modes and the time functions.



## 4.2 Single-field-based PC/EOF Analysis

In this method, the normalized time functions of one of the fields are determined using PC or EOF analysis. Then the spatial patterns of both the fields are determined, using the time functions already determined for the first data field. For the first data field, the first pair of time and space functions accounts for the maximum variance of the data field. The corresponding spatial pattern for the second field is evidently the pattern that correlates with the spatial pattern of the first field, which accounts for the maximum variance of the first field.

## 4.3 SVD of Covariance Matrix

For examining the coupled variability of two physically related time dependent data fields, the Singular Value Decomposition (SVD) of the cross-covariance matrix is for used by Bretherton et al. [1992] and Venegas et al. [1996]. From the two given data fields, this method determines pairs of orthonormal spatial vectors and their temporal variations, which explain the maximum variance of the covariance between the two fields.

Suppose we have one set of variables  $X$  at  $p$  stations and the other set  $Y$  at  $q$  stations, both observed at  $N$  time levels as defined in Chapter 4.1. We shall form the temporal cross-covariance matrix,

$$C_{xy} = X^T Y$$

where  $C_{xy}$  is matrix of size  $(p, q)$ . Note that, this cross-covariance matrix may not be a square matrix, as the two fields may be defined over different grids ( $p \neq q$ ), but the variables need to be observed at the same number of time levels,  $N$ . We may also consider the observation on one of the variables to lag behind the other and examine the lagged cross-covariance structure for the possible inter-relationships. At this point it should be noted that some investigators choose to scale the variability in  $X$  and  $Y$ , by dividing each time series by its standard deviation. The reason behind this kind of scaling is that if one set of variables has much larger amplitude variability than the other, the variability in that field might dominate the covariance structure between the matrices.

Once the cross-covariance matrix  $C_{xy}$  has been formed, we perform the singular value decomposition of  $C_{xy}$ . That is  $C_{xy}$  is expressed in the form,

$$C_{xy} = U \Gamma V^T$$

where the columns of  $U$  and  $V$ , the eigenvectors of the cross-product matrices  $C_{xy} C_{xy}^T$  and  $C_{xy}^T C_{xy}$ , are the orthonormal basis vectors in the  $p$ - and  $q$ -dimensional variable spaces respectively. From

the above equation, it should also be evident that  $U$  and  $V$  describe the spatial modes of co-variability between the fields  $X$  and  $Y$ . Hence,  $U$  and  $V$  can be considered to describe the orthonormal coupled spatial patterns/modes of the variables  $X$  and  $Y$  respectively. Just as with the EOF, these modes, which could be termed as the SVD modes of the data fields  $X$  and  $Y$ , represent standing oscillations in the data fields.

Each pair of the column vectors of  $U$  and  $V$  is associated with each diagonal element of  $\Gamma$ , where  $\Gamma$  is a diagonal matrix with the square roots of the eigenvalues of the cross-product matrix  $C_{xy}^T C_{xy}$ , arranged in the descending order of magnitude along its diagonal. Hence, the total squared covariance of  $C_{xy}$  is given by the sum of the squared diagonal values of  $\Gamma$ . This gives a simple way of assessing the relative importance of the eigenmodes, through the squared covariance (SC) explained by each mode. If  $\gamma_i = \Gamma(i, i)$  is the  $i$ -th singular value, the fraction of squared co-variance (SCF) explained by the eigenvector  $U_i$  and  $V_i$  is given by,

$$SCF_i = \gamma_i^2 / (\sum \gamma_i^2)$$

As the singular values of  $C_{xy}$  are arranged in the descending order of magnitude, along the diagonal of the matrix  $\Gamma$ , the first pair of column vectors of  $U$  and  $V$  explains the largest fraction of  $SC$ , and each succeeding pair describes a maximum fraction of the variance of  $SC$  that is unexplained by previous pairs.

We can now find the time series describing how each mode of variability oscillates in time. For the given  $U$ , the time function of  $X$  is given by,

$$\xi = XU$$

and for  $Y$  we have,

$$\eta = YV$$

The column vectors of the matrices  $\xi$  and  $\eta$  contain the time function of the fields corresponding to the spatial modes  $U$  and  $V$ . Note that the column vectors of  $\xi$  and  $\eta$  may not be orthogonal. Also note that using a few columns of the spatial patterns and time functions, we can get the optimal estimates of the data matrices; i.e.,  $X \approx \hat{X} = \xi U^T$  and  $Y \approx \hat{Y} = \eta V^T$ .

## 4.4 Canonical Correlation Analysis

In this section we briefly review the Canonical Correlation Analysis (CCA), also referred to as Canonical Variate Analysis by Pielou [1977]. This method is designed to identify the linear combinations of variables in one field that are most strongly correlated with linear combinations of

variables in another field. For example, Barnett [1981] and Barnett and Preisendorfer [1987] have used CCA to relate patterns of seasonal mean SST anomalies over the Pacific to surface-air temperature anomalies over the United States the following season.

#### 4.4.1 Description of Canonical Variables

In CCA, we are concerned with the interdependence between two sets/groups of variables, which are physically related and measured at number of time levels. Suppose we have one set of variables  $X$  at  $p$  stations and the other set  $Y$  at  $q$  stations ( $p \leq q$ ), both observed at  $N$  time levels forming a combined data matrix, denoted by  $[X, Y]$  of size  $(N, p + q)$ , as defined in Chapter 4.1.

We now seek two new sets of variables with  $p$  elements in each set,  $[\xi_1, \xi_2, \dots, \xi_p]$  and  $[\eta_1, \eta_2, \dots, \eta_p]$ , where each  $\xi_i$  and  $\eta_i$  represent a column vector of  $N$  elements constructed through a linear transformation of the original variables such that, the correlation matrix  $R$  of the new variables  $[\xi, \eta]$  has the form,

$$R = \left[ \begin{array}{cccc|cccc} 1 & 0 & \dots & 0 & r_1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & r_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & r_p \\ \hline r_1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & r_2 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & r_p & 0 & 0 & \dots & 1 \end{array} \right]$$

The new variables  $[\xi_1, \xi_2, \dots, \xi_p]$  and also  $[\eta_1, \eta_2, \dots, \eta_p]$  are known as the canonical variates/variables.

#### 4.4.2 Determination of Canonical Variables

We will discuss here the mathematical manipulation involved in the determination of these canonical variables. This problem is approached by first looking for a stationary correlation between one member from  $\xi$ -group and one member from  $\eta$ -group, and then proving that the chosen member of  $\xi$ -group is uncorrelated with other members of both  $\xi$ - and  $\eta$ - groups. Note that the above property that we seek between the member of  $\xi$ - and  $\eta$ - groups is possible only when  $N \geq (2p - 1)$ .

Consider the first pair of canonical variables  $\xi_1$  and  $\eta_1$  formed from the given data matrix through a linear transformation defined in the form,

$$\begin{aligned}\xi_1 &= a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p \\ \eta_1 &= b_{11}Y_1 + b_{21}Y_2 + \dots + b_{q1}Y_q\end{aligned}$$

where  $X_i$  and  $Y_i$  are the column vectors of the data matrix  $[X, Y]$ , and

$$\begin{aligned}a_1 &= [a_{11}, a_{21}, \dots, a_{p1}]^T \\ b_1 &= [b_{11}, b_{21}, \dots, b_{q1}]^T\end{aligned}$$

are column vectors of the transformation coefficient matrices  $a$  and  $b$  of size  $(p, p)$  and  $(q, p)$  respectively. The superscript  $T$  denotes the transpose of a matrix or a vector. The second subscripts of  $a$  and  $b$  in the above equation refer to the canonical variable considered.

Then the variance of  $\xi_1$  is given by,

$$\text{Var}[\xi_1] = \text{Var}[a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p] = a_1^T C_{xx} a_1$$

where  $C_{xx}$  is the covariance matrix of order  $(p, p)$ , whose elements are given by,

$$C_{xx}(i, j) = C_{xx}(j, i) = \text{Cov}(X_i, X_j)$$

where  $\text{Cov}(\cdot)$  stands for the covariance defined as,

$$\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i) = (1/N) \sum_{k=1}^N (x_{ki} - \bar{X}_i)(x_{kj} - \bar{X}_j)$$

and  $\bar{X}_i, \bar{X}_j$  are the arithmetic means of the  $i$ -th and  $j$ -th columns of the data matrix  $X$  respectively. Similarly we have,

$$\text{Var}[\eta_1] = b_1^T C_{yy} b_1$$

where  $C_{yy}$  is the covariance matrix of order  $(q, q)$  whose elements are given by,

$$C_{yy}(i, j) = C_{yy}(j, i) = \text{Cov}(Y_i, Y_j)$$

The covariance between  $\xi_1$  and  $\eta_1$  is given by,

$$\text{Cov}(\xi_1, \eta_1) = a_1^T C_{xy} b_1 = b_1^T C_{yx} a_1$$

where the covariance matrices  $C_{xy}$  and  $C_{yx}$  are of order  $(p, q)$  and  $(q, p)$  respectively, with  $C_{yx} = C_{xy}^T$  and are defined as follows:

$$C_{xy}(i, j) = Cov(X_i, Y_j) = Cov(Y_j, X_i) = C_{yx}(j, i)$$

Now for  $\xi_1$  and  $\eta_1$  to form a pair of canonical variables, we require the correlation coefficient between them defined as,

$$\begin{aligned} r_1(\xi_1, \eta_1) &= Cov(\xi_1, \eta_1) / [Var(\xi_1)Var(\eta_1)]^{1/2} \\ &= [a_1^T C_{xy} b_1] / D = [b_1^T C_{yx} a_1] / D \end{aligned}$$

to be stationary for variation in the transformation coefficient vectors  $a$  and  $b$ , where  $D = [a_1^T C_{xx} a_1 \cdot b_1^T C_{yy} b_1]^{1/2}$ , subject to the constraints that the canonical variables have unit variance, i.e.,

$$\begin{aligned} Var[\xi_1] &= a_1^T C_{xx} a_1 = 1 \\ Var[\eta_1] &= b_1^T C_{yy} b_1 = 1 \end{aligned}$$

which only mean that we are looking for normalized canonical variables. That is, the problem can be defined as the maximization of the function,

$$r_1(\xi_1, \eta_1) = [a_1^T C_{xy} b_1] = [b_1^T C_{yx} a_1]$$

subject to the constraints  $a_1^T C_{xx} a_1 = 1$  and  $b_1^T C_{yy} b_1 = 1$ . This maximization problem is solved by using the method of Lagrange's undetermined multipliers, which leads to the determination of the unconditional extremum of the function,

$$f(a_1, b_1) = [a_1^T C_{xy} b_1] - (\lambda/2)[a_1^T C_{xx} a_1 - 1] - (\mu/2)[b_1^T C_{yy} b_1 - 1]$$

where  $\lambda$  and  $\mu$  are undetermined multipliers.

For  $f(a_1, b_1)$  to be extremum we require,

$$\frac{\partial f(a_1, b_1)}{\partial a_1} = C_{xy} b_1 - \lambda C_{xx} a_1 = 0 \quad (4.1)$$

$$\frac{\partial f(a_1, b_1)}{\partial b_1} = C_{yx} a_1 - \mu C_{yy} b_1 = 0 \quad (4.2)$$

Premultiplying (4.1) by  $a_1^T$  and (4.2) by  $b_1^T$ , and subtracting one from the other, we get

$$r(\xi_1, \eta_1) = \lambda = \mu$$

Now premultiplying (4.2) by  $C_{xy}C_{yy}^{-1}$ , substituting  $\lambda C_{xx}a_1$  for  $C_{xy}b_1$  from (4.1) and noting that  $\lambda = \mu$ , we get,

$$(C_{xx}^{-1}C_{xy}C_{yy}^{-1}C_{xy}^T - \lambda^2 I_p)a_1 = 0 \quad (4.3)$$

where  $I_p$  is a unit matrix of order  $(p, p)$ . (4.3) could be easily recognized to define a matrix eigenvalue problem. The matrix  $C_{xx}^{-1}C_{xy}C_{yy}^{-1}C_{xy}^T$ , being of order  $(p, p)$ , has  $p$  eigenvalues. Now the eigenvalues could be computed along with the eigenvectors using standard procedures. The eigenvectors corresponding to different eigenvalues defined by (4.3) form the column vectors of the transformation coefficient matrix  $a$ .

Starting from (4.1) and following the procedure similar to that outlined for the determination of the transformation coefficient matrix  $a$ , it can be shown that the matrix  $b_1$  satisfies the relation,

$$(C_{yy}^{-1}C_{xy}^T C_{xx}^{-1}C_{xy} - \lambda^2 I_q)b_1 = 0 \quad (4.4)$$

where  $I_q$  is unit matrix of order  $(q, q)$ . The matrix pre-multiplying  $b_1$  in (4.4) is of order  $(q, q)$  and obviously the problem defined by (4.4) has  $q$  eigenvalues.

We shall show here that the non-zero eigenvalues of the problems defined by (4.3) and (4.4) are identical. Pre-multiplying (4.4) by  $C_{xx}^{-1}C_{xy}$ , we get,

$$(C_{xx}^{-1}C_{xy}C_{yy}^{-1}C_{xy}^T - \lambda^2 I_p)C_{xx}^{-1}C_{xy}b_1 = 0 \quad (4.5)$$

It can be easily seen that the unit matrix  $I_p$  is of order  $(p, p)$  and that the eigenvalue problem defined by (4.5) is same as that defined by (4.3). In other words all the non-zero eigenvalues of (4.4) are also the eigenvalues of the problem defined by (4.3). It should be further noted that since the number of eigenvalues of the problem defined by (4.5) couldn't be greater than  $p$ .

#### 4.4.3 Correlation between $\xi_i$ and $\eta_j$

Finally it remains to be shown that  $\xi_i$  and  $\eta_j$  are uncorrelated between and among themselves when  $i \neq j$ . For any  $i$ , we have from (4.1) and (4.2),

$$\sum_{n=1}^q C_{xy}(m, n)b_{ni} = \lambda_i \sum_{n=1}^p C_{xx}(m, n)a_{ni} \quad (4.6)$$

$$\sum_{n=1}^p C_{yx}(m, n)a_{ni} = \lambda_i \sum_{n=1}^q C_{yy}(m, n)b_{ni} \quad (4.7)$$

There will be a similar pair for the other suffix  $j$ . Among the four variables  $\xi_i, \xi_j, \eta_i,$  and  $\eta_j$  there are six covariances and it is now required to be shown that four of them vanish, namely,

$$\begin{aligned} Cov(\xi_i, \xi_j) &= \sum_{m=1}^p \sum_{n=1}^p a_{ni} a_{mj} Cov(x_i, x_j) \\ &= \sum_{m=1}^p \sum_{n=1}^p a_{ni} a_{mj} C_{xx}(m, n) \end{aligned} \quad (4.8)$$

$$Cov(\xi_j, \eta_i) = \sum_{m=1}^p \sum_{n=1}^q a_{mj} b_{ni} C_{xy}(m, n) \quad (4.9)$$

$$Cov(\xi_i, \eta_j) = \sum_{m=1}^q \sum_{n=1}^p a_{ni} b_{mj} C_{yx}(m, n) \quad (4.10)$$

$$Cov(\eta_i, \eta_j) = \sum_{m=1}^q \sum_{n=1}^q b_{ni} b_{mj} C_{yy}(m, n) \quad (4.11)$$

Multiply (4.6) by  $a_{mj}$  and summing over  $m$ , we find by virtue of (4.8) and (4.9),

$$Cov(\xi_j, \eta_i) = \lambda_i Cov(\xi_i, \xi_j) \quad (4.12)$$

Similarly multiplying (4.7) by  $b_{mj}$  and summing over  $m$  we find,

$$Cov(\xi_i, \eta_j) = \lambda_j Cov(\eta_i, \eta_j) \quad (4.13)$$

Inter changing  $i$  and  $j$  in (4.12) and using (4.13), we get,

$$\lambda_j Cov(\xi_i, \xi_j) = \lambda_i Cov(\eta_i, \eta_j) \quad (4.14)$$

and interchanging  $i$  and  $j$  in this,

$$\lambda_i Cov(\xi_j, \xi_i) = \lambda_j Cov(\eta_i, \eta_j) \quad (4.15)$$

It follows from (4.14) and (4.15) that either

$$Cov(\xi_i, \xi_j) = Cov(\eta_i, \eta_j) = 0 \quad (4.16)$$

or

$$\lambda_i^2 = \lambda_j^2 \quad (4.17)$$

But, since  $\lambda_i$  and  $\lambda_j$  are two distinct eigenvalues of the problem defined by (4.3), only the relation (4.16) is true. In a similar manner it may be shown that the other covariances also vanish.

#### 4.4.4 Spatial Patterns and Time functions

The normalized column vectors of the transformation coefficient matrices  $a$  and  $b$  are respectively the basis vectors in  $p$ - and  $q$ -dimensional spaces of the variables  $X$  and  $Y$ . These orthonormal basis vectors can be considered to describe the spatial patterns/modes of the variables and represent

standing oscillations in the data fields  $X$  and  $Y$ . Note that, each pair of the column vectors of matrices  $a$  and  $b$  describes spatial modes, which are strongly coupled; i.e. with maximum correlation between the corresponding pairs of time functions. The functions describing how each mode of variability oscillates in time are given by  $\xi$  and  $\eta$ , and the scalar products between the column vectors of the time functions provide a measure of the coupling between different spatial modes of the data fields.

#### 4.4.5 Limitations

In geophysical applications of CCA, when the number of observation time is not much larger than the number of variables (i.e. the sampling grid points), we are likely to get spurious correlations between the derived canonical variates. Also sometimes the number of observation times may be less than the number of grid points in the spatial fields, and in this case we can not directly construct canonical variates since the matrices  $C_{xx}$  and  $C_{yy}$  are not invertible. To circumvent these limitations of CCA Barnett and Preisendorfer [1987] have suggested an alternative approach.

### 4.5 Barnett and Preisendorfer method

In this method, the number of variables is reduced by taking only a few leading time functions of EOF analysis of the original data fields (i.e. the contributions from higher time functions are filtered), and then SVD is applied to the covariance matrix between the reduced data fields.

Let  $X$  and  $Y$  be the two data fields of size  $(N, p)$  and  $(N, q)$  respectively. Using the EOF analysis, we may write,

$$X = U_x \Gamma_x V_x^T \text{ and } Y = U_y \Gamma_y V_y^T$$

where the column vectors of  $V_x$  (of dimension  $p$ ) and  $V_y$  (of dimension  $q$ ) are the orthonormal spatial functions of the EOF expansion of the data fields  $X$  and  $Y$  respectively. The columns vectors of  $U_x$  and  $U_y$  are the orthonormal time functions and  $\Gamma_x$  and  $\Gamma_y$  contain along their diagonal the singular values of  $X$  and  $Y$  respectively.

We will consider only a few leading time functions of the data fields  $X$  and  $Y$ , and form the cross-covariance matrix,

$$C_{mn} = \hat{U}_x^T \hat{U}_y$$

where  $\hat{U}_x$  and  $\hat{U}_y$  contain only the first few leading vectors of  $U_x$  and  $U_y$  (say the first  $m$  vectors of  $U_x$  and  $n$  vectors of  $U_y$ ) such that the truncated EOF expansions represent a specified percentage of the variance of the original data fields. In forming the above cross covariance matrix  $C_{mn}$ ,  $\hat{U}_x$



and  $\hat{U}_y$  are treated as data fields though they are actually the time functions of the fields  $X$  and  $Y$ . Then the SVD of  $C_{mn}$  yields,

$$C_{mn} = \alpha_m \Gamma \beta_n^T$$

where the column vectors of  $\alpha_m$  (of dimension  $m$ ) and  $\beta_n$  (of dimension  $n$ ) are the SVD modes, which are now the spatial modes. They are orthonormal in the reduced data fields of spatial dimensions  $m$  and  $n$  defined by considering only a few leading time functions/vectors in the EOF expansion of the data fields  $X$  and  $Y$ . These SVD modes account for most of the covariance between the two fields  $\hat{U}_x$  and  $\hat{U}_y$ , which represent a specified percentage of the variance of the original data fields  $X$  and  $Y$ .

The time functions corresponding to  $\alpha_m$  and  $\beta_n$  are given by,

$$\xi(t) = \hat{U}_x \alpha_m \text{ and } \eta(t) = \hat{U}_y \beta_n$$

where  $\hat{U}_x$  and  $\hat{U}_y$  are treated as data fields. Note that the column vectors of  $\xi$  are orthonormal vectors in an  $N$ -dimensional space, i.e.,

$$\xi^T \xi = \alpha_m^T \hat{U}_x^T \hat{U}_x \alpha_m = I$$

since  $\hat{U}_x$  and  $\alpha_m$  are orthonormal vectors. Similarly we can show that the column vectors of  $\eta$  are orthonormal vectors.

Once  $\xi$  and  $\eta$  are determined, the corresponding spatial patterns of the  $X$  and  $Y$  fields are defined respectively as,

$$V_p = X^T \xi \text{ and } V_q = Y^T \eta \quad (4.18)$$

where the column vectors of  $V_p$  and  $V_q$  are of dimensions  $p$  and  $q$  respectively, and define the spatial modes of the data fields  $X$  and  $Y$ .

Note that, as the time functions  $\xi$  and  $\eta$  are constructed using the SVD modes  $\alpha_m$  and  $\beta_n$ , (which account for most of the covariance between the two fields  $\hat{U}_x$  and  $\hat{U}_y$ , that represent a specified percentage of the variance of the original data fields  $X$  and  $Y$ ) these time functions will explain essentially most of the covariance between the original data fields  $X$  and  $Y$ . Consequently vide (4.18) each pair of the column vectors of  $V_p$  and  $V_q$  describes spatial modes, which are strongly coupled in the sense discussed above.