

## MULTIVARIATE DATA ANALYSIS

A.A. Fernandes, M.K. Antony, Y.K. Somayajulu  
Y.V.B. Sarma, A.M. Almeida

Physical Oceanography Division  
**National Institute of Oceanography**

Dona Paula  
Goa 403 004  
INDIA

---



---

R. Mahadevan  
Visiting Scientist  
National Institute of Oceanography, Goa, INDIA  
Centre for Earth Science Studies, Thiruvananthapuram, Kerala, INDIA

Technical Report No. NIO/TR-10/2005  
October 2005

For any enquiries regarding the contents of this technical report, including the software listed, please contact:

Dr. A. A. Fernandes  
Scientist  
Physical Oceanography Division  
National Institute of Oceanography  
Dona Paula, Goa  
INDIA 403 004

E-mail: [fernan@darya.nio.org](mailto:fernan@darya.nio.org)

Phone: +91-832-2450296

FAX : +91-832-2450602 & 2450603

---

# Our Publications on Multivariate & time series data analysis

- *Factor analysis/ EOF analysis*: Fernandes and Mahadevan [1982], Mahadevan and Fernandes [1985]
- *Statistical prediction of paleoclimate*: Fernandes and Gupta [1994], Gupta and Fernandes [1995], Gupta and Fernandes [1997], Gupta et al. [1996]
- *Generalized Inverse*: Menezes and Mahadevan [1984], Murty and Mahadevan [1995], Fernandes et al. [2002]
- *Spectrum & Cross spectrum analysis*: Fernandes et al. [1981], Gouveia and Mahadevan [1983], Fernandes et al. [1988], Fernandes [1999], Kumar and Fernandes [1994], Fernandes et al. [2000a], Fernandes et al. [2000b], Fernandes et al. [2002],
- *Rotary spectrum analysis*: Fernandes et al. [1993], Fernandes and Pednekar [1997]
- *Low pass & Band pass filters*: Fernandes et al. [1983], Fernandes et al. [1994]
- *Harmonic analysis*: Fernandes et al. [1991], Fernandes et al. [1993]
- *Bispectrum analysis*: Sengupta and Mahadevan [1983]
- *Data assimilation*: Mahadevan et al. [2001]
- *Complex EOF analysis & Canonical Correlation analysis*: present work
- *Work in progress*: Computation of absolute ocean currents from temperature, salinity data using the generalized inverse method (geostrophy gives only relative currents, relative to an **assumed** level of no motion).
- *Work to be taken up shortly*: Statistical prediction of climate including El Nino/ La Nina events, using Canonical Correlation analysis. The importance of this work lies in the fact that, except for the 1997-98 event, El Nino events have coincided with deficient south west monsoon rainfall over India, resulting in drought; while La Nina events have been associated with excessive rainfall, resulting in floods.

# Acknowledgements

At the outset we wish to express our gratitude to Dr. Ehrlich Desa, Ex-Director, National Institute of Oceanography (NIO), Dona Paula, for providing financial support from Lab Reserves for the visits of Dr. R. Mahadevan, Retired Professor, Indian Institute of Technology, Madras, to NIO Goa. This report constitutes part of the work carried out by scientists from NIO with the collaboration and guidance of Dr. Mahadevan.

I wish to express my heart felt gratitude to Dr. J. O'Brien for helpful discussions via e-mail, which helped clarify my doubts concerning computations of phase speed of Rossby and Kelvin waves from spatial and temporal phase functions derived from Complex EOF analysis. I am greatly indebted to Dr. P. N. Vinayachandran, who is one of the co-authors in the seminal work on Indian Ocean dipole [Saji et al., 1999] for confirming that our finding of Indian Ocean Dipole from results of Complex EOF analysis of TOPEX/Poseidon data in the Indian Ocean was indeed correct, and also for bringing to our attention the work of Chambers et al. [1999], who reported the signature of Indian Ocean dipole using EOF analysis.

I wish to thank Mr. M. Babu for bringing to my attention the work of Eigenheer and Quadfasel [2000], who analyzed TOPEX/Poseidon sea level height anomaly data in the Bay of Bengal using Complex EOF analysis. I wish to thank Mr. O. P. Sreejith for providing global *Reynolds* SST data in ASCII form. I wish to thank Dr. D. Shankar for helping in interpretation of our results of Complex EOF analysis of Reynolds SST data in the Bay of Bengal as well as for cautioning me about the futility of my venture in the field of Complex EOF analysis, since all the relevant data for the Indian Ocean has already been analysed using this technique and the results published in reputed journals! I wish to thank Dr. Prasanna Kumar and Mr. Nuncio Murukesh, who kindly provided the address of the web site of Center for Space Research, University of Texas, Austin for downloading the global TOPEX/Poseidon sea level height anomaly data. We are grateful to Dr. D.P. Chambers of the Center for Space Research, for enabling us to download the revised and upto date TOPEX/Poseidon data used in this report.

Dr. T. Pankajakshan kindly provided a copy of an unpublished manuscript "Climatic variability of sea surface temperature and wind field over Indian Ocean" by T. Pankajakshan (in association with R. Mahadevan).

I wish to express my deep gratitude to Dr. M. R. Rameshkumar, who provided the funds from award monies earned by him, for enabling me to present a paper “Complex EOF analysis of sea level anomaly in the Indian Ocean at ENSO time scale” at the *International Workshop on the Role of the Indian Ocean in Climate Variability over India (INDOCLIM)* held at IITM, Pune during 23-27 February 2004. I am also indebted to Dr. Rameshkumar for recently providing me with a suitable personal computer, without which this report would not have seen the light of day. I wish to thank Dr. S. S. C. Shenoi and Dr. P. Vethamony for allowing me access to the Sun work stations, Sahyadri and Sarayu respectively.

I wish to place on record my sincere thanks to Prof. Ashish Sengupta, Head Applied Statistics Unit, Indian Statistical Institute, Calcutta for going through the section on Canonical Correlation Analysis and offering his comments on the same.

This report has been prepared using “Latex” on a “Linux” platform, viz., the personal computer *Kapila*. I wish to thank Mr. Dattaram Shivji for installing “Latex” and “GMT” packages on the personal computer. The style file used for preparing this report, has been hacked by me from a Goa University, Ph. D style file prepared by Dr. D. Shankar, to whom I am indebted for teaching me the art of *word processing* in “Latex”. The colour *Hov-moller* diagrams as well as TOPEX/Poseidon spatial EOF and Complex EOF derived fields were constructed using a specimen *GMT* code kindly provided by Mr. Michael Selvam.

Ms. Felicidade Alfonso provided the secretarial assistance.

The authors wish to thank the Director NIO for the facilities provided.

A. A. FERNANDES

*National Institute of Oceanography, Goa*  
*October 2005*

# Abstract

This report describes some of the multivariate data analysis techniques used in meteorology and oceanography and provides the necessary background and software to undertake appropriate analyses of a given data set. The multivariate data analysis techniques discussed are for the identification of simple patterns within a data set and between data sets, and for studying the characteristics of moving patterns in space-time data sets.

Among the data analysis techniques discussed are Empirical Orthogonal Function (EOF) Analysis, Complex EOF Analysis, and Canonical Correlation Analysis, which are useful in the description and prediction of climate.

Results of EOF and Complex EOF analysis of TOPEX/Poseidon derived sea level height anomaly data in the Indian and Pacific Oceans at inter-annual time scales are presented. Results of Canonical Correlation Analysis of a sample data set is also presented.

# Contents

	<b>i</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>Abstract</b>	<b>vi</b>
<b>List of Tables</b>	<b>x</b>
<b>List of Figures</b>	<b>xi</b>
<b>1 INTRODUCTION</b>	<b>1</b>
1.1 Scope of the report . . . . .	3
<b>2 PATTERNS WITHIN MULTIVARIATE DATA</b>	<b>4</b>
2.1 Singular Value Decomposition . . . . .	4
2.2 Principal Component Analysis . . . . .	7
2.2.1 Stationary Vector . . . . .	7
2.3 Empirical Orthogonal Function Analysis . . . . .	9
2.3.1 Interpretation of EOF . . . . .	10
2.3.2 Advantages . . . . .	11
2.3.3 Shortcomings . . . . .	11
2.3.4 EOF Analysis and Statistical Forecasting . . . . .	13
2.3.5 Statistical and Dynamic methods of prediction . . . . .	14
2.3.6 Variations in conventional EOF . . . . .	14
2.4 Factor Analysis . . . . .	16
2.5 Varimax Rotation . . . . .	17
2.5.1 Rotation of Principal components with complex elements . . . . .	21
2.6 Number of Components to retain . . . . .	21
2.7 Correlation matrix or covariance matrix? . . . . .	22

<b>3</b>	<b>Moving patterns in space-time data sets</b>	<b>24</b>
3.1	Analysis of Fourier coefficients of time series . . . . .	24
3.2	Cross spectral analysis . . . . .	25
3.3	Lagged cross correlation analysis . . . . .	26
3.4	Frequency Domain Complex PCA (FDPCA) . . . . .	26
3.5	Time Domain Complex PCA (TDCPCA) . . . . .	27
<b>4</b>	<b>Relationships between data sets</b>	<b>30</b>
4.1	Combined Principal Component Analysis . . . . .	31
4.2	Single-field-based PC/EOF Analysis . . . . .	32
4.3	SVD of Covariance Matrix . . . . .	32
4.4	Canonical Correlation Analysis . . . . .	33
4.4.1	Description of Canonical Variables . . . . .	34
4.4.2	Determination of Canonical Variables . . . . .	34
4.4.3	Correlation between $\xi_i$ and $\eta_j$ . . . . .	37
4.4.4	Spatial Patterns and Time functions . . . . .	38
4.4.5	Limitations . . . . .	39
4.5	Barnett and Preisendorfer method . . . . .	39
<b>5</b>	<b>Examples of Horel [1984]</b>	<b>41</b>
5.1	Complex EOF Analysis: Computer Simulated data . . . . .	41
5.1.1	Example 1 . . . . .	41
5.1.2	Example 2 . . . . .	46
5.1.3	Example 3 . . . . .	46
5.1.4	Example 4# . . . . .	47
<b>6</b>	<b>TOPEX/Poseidon Sea Level Height Anomaly</b>	<b>61</b>
6.1	Complex EOF Analysis: TOPEX/Poseidon - Indian & Pacific Oceans . . . . .	62
6.2	EOF Analysis: TOPEX/Poseidon - Indian & Pacific Oceans . . . . .	70
<b>7</b>	<b>Summary</b>	<b>80</b>
<b>A</b>	<b>Software</b>	<b>81</b>
A.1	EOF Analysis . . . . .	82
A.1.1	Computer Program: eof-vec.f . . . . .	82
A.1.2	Specification File: eof-vec.spc . . . . .	104
A.2	Hilbert Transform . . . . .	105
A.2.1	Computer Program: hilbert.f . . . . .	105



---

A.2.2	Specification file: hilbert.spc . . . . .	119
A.3	Complex EOF Analysis . . . . .	120
A.3.1	Computer Program: ceof-vec.f . . . . .	120
A.3.2	Specification File: ceof-cpc.spc . . . . .	150
A.4	Cross Spectrum Analysis . . . . .	151
A.4.1	Computer Program: cross-spec.f . . . . .	151
A.4.2	Specification File: cross-spec.spc . . . . .	172
A.4.3	Correction . . . . .	172
A.5	Canonical Correlation Annalysis . . . . .	173
A.5.1	Computer Program: canonical.f . . . . .	173
A.5.2	Data: pielou-page-357.dat . . . . .	190
A.5.3	Results: see-cca.res . . . . .	191
<b>Bibliography</b>		<b>194</b>

# List of Tables

2.1	Significance test for selecting number of EOFs to retain . . . . .	23
6.1	T-P data - Indian Ocean (Interannual): CEOF - Variance explained . . . . .	63
6.2	T-P data - Pacific Ocean (interannual): CEOF - Variance explained . . . . .	63
6.3	T-P data - Indian Ocean (Interannual): EOF - Variance explained . . . . .	70
6.4	T-P data - Pacific Ocean (interannual): EOF -variance explained . . . . .	70

# List of Figures

2.1	Singular Value Decomposition of a Matrix $X$ . . . . .	5
2.2	Principal Component Analysis . . . . .	8
5.1	Example 1: Data & Hilbert transform . . . . .	42
5.2	Example 2: Data & Hilbert transform . . . . .	43
5.3	Example 3: Data & Hilbert transform . . . . .	44
5.4	Example 4#: Data & Hilbert transform . . . . .	45
5.5	Complex EOF analysis: Example 1. <i>Stick diagram</i> . . . . .	48
5.6	Complex EOF analysis: Example 1. <i>Amplitude &amp; phase</i> . . . . .	49
5.7	Complex EOF analysis: Example 2. <i>Stick diagram</i> . . . . .	50
5.8	Complex EOF analysis: Example 2. <i>Amplitude &amp; phase</i> . . . . .	51
5.9	Complex EOF analysis: Example 3. <i>Stick diagram</i> . . . . .	52
5.10	Complex EOF analysis: Example 3. <i>Amplitude &amp; phase</i> . . . . .	53
5.11	Complex EOF analysis: Example 4#. <i>Stick diagram</i> . . . . .	54
5.12	Complex EOF analysis: Example 4#. <i>Amplitude &amp; phase</i> . . . . .	55
5.13	Example 1: Data & Reconstructed data from CEOF-1 . . . . .	56
5.14	Example 2: Data & Reconstructed data from CEOF 1 & 2 . . . . .	57
5.15	Example 3: Data & Reconstructed data from CEOF 1 & 2 . . . . .	58
5.16	Example 4#: Data & Reconstructed data from CEOF 1 . . . . .	59
5.17	Example 4#: Data & Reconstructed data from CEOF 1 & 2 . . . . .	60
6.1	Low Pass Filter Response functions . . . . .	64
6.2	T-P data - Indian Ocean (Interannual): Spatial CEOF-1 Amplitude . . . . .	65
6.3	T-P data - Indian Ocean (Interannual): Spatial CEOF-1 Phase . . . . .	66
6.4	T-P data - Indian Ocean (Interannual): Temporal CEOF 1&2 . . . . .	67
6.5	T-P data - Pacific Ocean (Interannual): Spatial CEOF-1 Amplitude . . . . .	68
6.6	T-P data - Pacific Ocean (Interannual): Temporal CEOF 1&2 . . . . .	69
6.7	T-P data - Indian Ocean (Interannual): CEOF-1 Hov-moller Rossby waves . . . . .	71

---

6.8	T-P data - Indian Ocean (interannual): CEOF-1 Hov-Moller Kelvin wave . . . . .	72
6.9	T-P data - Indian Ocean (annual): CEOF-1 Hov-Moller Rossby waves . . . . .	73
6.10	T-P data - Indian Ocean (annual): CEOF-1 Hov-Moller Kelvin wave . . . . .	74
6.11	T-P data - Indian Ocean (Interannual): spatial EOF-1 elevation . . . . .	75
6.12	T-P data - Indian Ocean (Interannual): Temporal EOF 1&2 . . . . .	76
6.13	T-P data - Pacific Ocean (Interannual): spatial EOF-1 elevation . . . . .	77
6.14	T-P data - Pacific Ocean (Interannual): Temporal EOF 1&2 . . . . .	78
6.15	T-P data - Pacific Ocean (Interannual): spatial EOF-2 elevation . . . . .	79

# Chapter 1

## INTRODUCTION

In meteorology and oceanography we often come across multivariate data sets of large dimensions, which are generally formed from repeated observations on several physical variables. These variables may be continuous over space and time, but observations on them are generally made only at discrete points, which may not necessarily be over an uniformly spaced grid in space. For example, the sea surface temperature (SST) measured at synoptic hours at several stations over an ocean surface (the Indian Ocean) for several years forms a multivariate data set. Physical variables such as SST are continuous in both space and time. But due to the practical limitation in the observational system we can collect data only at limited number of spatial points and time steps. Hence the given data set may be considered as a sample from a much larger continuously varying space-time function/field of the variable. Such large data sets of physical variables are analyzed to

- summarize and extract salient features embedded within them, and
- examine relationships between them.

Methods used for these analyses include estimation of simple spatial and temporal averages, autocorrelation and spectral density functions to sophisticated multivariate analyses. In this report some of the multivariate data analysis techniques used in meteorology and oceanography are discussed. Multivariate data can be considered as a set of points in a space of several dimensions. For example if the data set consists of 'N' samples on 'p' variables, it can be considered as 'N' points in the p-dimensional variable space or 'p' points in the N-dimensional sample space. If the dimension of the space in which the multivariate data is depicted is more than three, then picturing such a space in the mind becomes difficult. Hence different analysis techniques are devised to project the data points on to a space of reduced dimensions. Though the dimensions of the new data spaces may still be more than three, such analyses often provide better overview of important features of the phenomenon observed. All multivariate analyses generally aim at simplifying

and reducing the complexity of the problem. They examine whether there are structurally simpler ways of representing the complex data set under study, e.g. by transferring a set of interdependent variables to a set of independent variables or by reducing the dimensionality of the variable space. Consider, for example, a case in which we have  $p$  variables and are interested in their variances and correlations. There are  $p$  means,  $p$  variances and  $p(p-1)/2$  correlations (i.e.  $p(p+3)/2$  parameters altogether) to describe its variability. If we can transform the variables to an un-correlated set we can reduce the required parameters by  $p(p-1)/2$  (i.e. to  $2p$  parameters). Some of the analyses used to extract useful information from multivariate data sets are:

1. Analysis of patterns within a multivariate data
  - Principal Component (PC) Analysis
  - Empirical Orthogonal Function (EOF) Analysis
  - Factor Analysis
2. Analysis of moving patterns in space-time data
  - Analysis of Fourier coefficients of time series
  - Cross spectral analysis
  - Lagged cross correlation analysis
  - Frequency Domain PCA (FDPC)
  - Complex- PC and Complex-EOF Analyses
3. Analysis of relationships between data sets
  - Combined Principal Component Analysis (CPCA)
  - Single-field-based PCA (SFPCA)
  - SVD of Covariance matrix between two fields
  - Canonical Correlation Analysis (CCA)
  - Barnett and Preisendorfer method

The purpose of this report is to describe some of the multivariate data analysis techniques used in meteorology and oceanography and provide the necessary background and softwares to undertake appropriate analyses of a given data set.

## 1.1 Scope of the report

The multivariate data analysis techniques to identify simple patterns within a single data set are discussed in Chapter 2; techniques to detect moving patterns in space-time data set are discussed in Chapter 3; while in Chapter 4, the techniques to identify relationships between data sets are discussed.

The following softwares developed in FORTRAN for the multivariate data analysis and time series analysis are presented in the Appendix A:

- A software for the EOF analysis of a matrix with real elements, since most of the analyses of multivariate data sets are based on EOF analysis.
- Separate software for Cross Spectrum analysis, Low pass filter/ Band pass filter/ Hilbert transform, and Complex EOF analysis, required for Analyses of moving patterns in data sets.
- The software for Canonical Correlation Analysis to study relationships between data sets. Results of this analysis for “Test data” taken from Pielou [1977] are presented in this appendix. Canonical Correlation Analysis is one of the operational empirical methods used for the prediction of El Nino/ La Nina phenomena mentioned by Glantz [2001].

**In Chapter 5 the results of Complex EOF analysis of the synthetic data of Horel [1984] are presented. And in Chapter 6 Complex EOF analysis as well as EOF analysis of a typical geophysical data set, viz., the TOPEX/Poseidon derived sea level height anomaly data are presented.**

## Chapter 2

# PATTERNS WITHIN MULTIVARIATE DATA

Suppose that we have a multivariate data set formed by  $N$  samples of observations on  $p$  scalar variables. For example SST values at  $p$  stations (these stations are not necessarily over a uniformly spaced spatial grid) measured at times  $t_1, t_2, \dots, t_N$ . This multivariate data set can be considered as a matrix with  $N$ -rows and  $p$ -columns, and we will denote this data matrix as  $X$ . The entries in each column  $X_i$  of  $X$  form a time series of SST's at each station and can be considered as a position vector of a point in a  $N$ -dimensional sample space. The rows of  $X$  are called the maps or fields of the physical variable and can be considered to define  $N$  position vectors in a  $p$ -dimensional variable space. The statistical tools used to analyze such multivariate data sets of scalar fields are Principal Component Analysis (PCA), Empirical Orthogonal Function (EOF) analysis and the Factor analysis. These analyses are essentially the same except for some minor differences and can be explained through Singular Value Decomposition (SVD) of matrices. These analyses reduce the variability in the data set into a few patterns.

### 2.1 Singular Value Decomposition

We shall consider here the decomposition of a given matrix  $X$  of size  $(N, p)$  as a product of three matrices; two of them ( $U$  and  $V$ ) have orthonormal column vectors and the third  $\Gamma$  is a diagonal matrix. That is,

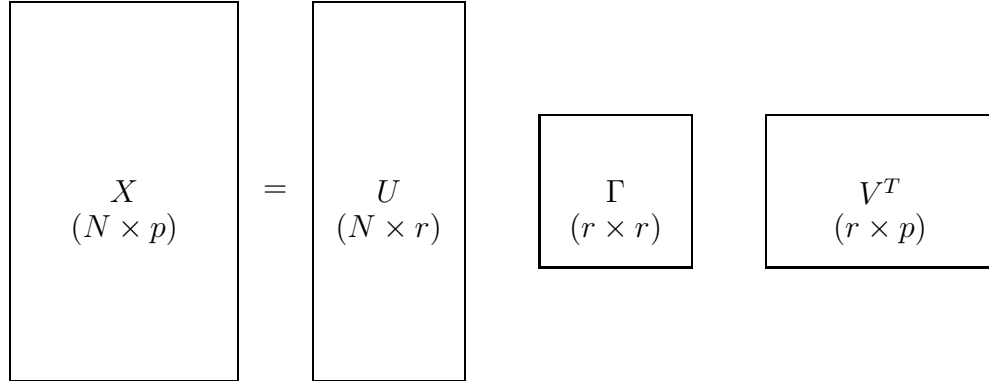
$$X = U\Gamma V^T \tag{2.1}$$

which is called the singular value decomposition of matrix  $X$ . Matrix  $U$  is of size  $(N, r)$  and  $V$  is of size  $(p, r)$  and  $r$  ( $r \leq \text{Min}\{N, p\}$ ) is the rank of the matrix  $X$ , where  $\text{Min}\{\cdot\}$  stands for minimum



value.  $\Gamma$  is an  $(r, r)$  diagonal matrix with diagonal elements  $\gamma_i$  called the singular values of  $X$ . The matrix multiplication involved in (2.1) is illustrated in Figure 2.1.

**Figure 2.1** Singular Value Decomposition of a Matrix  $X$



The rank of a matrix is defined as the dimension of the largest square sub-matrix of the given matrix whose determinant is non-zero. From this definition it follows that a matrix must have its rank equal to or less than the number of columns or rows in it, whichever is minimum. It also implies that the rank of a product matrix cannot exceed the rank of its components. Further, if the rank of a matrix is equal to the smaller dimension of a matrix, it cannot be split into component matrices having smaller order. By order of a matrix we refer here to the smaller dimension of the matrix. Obviously the ranks of all the three component matrices in (2.1) will be  $r$  and the diagonal elements of  $\Gamma$  are non-zero. The existence of such decomposition is *a priori* assumed in this analysis. We will discuss below the technique followed to generate  $U$ ,  $\Gamma$  and  $V$  in (2.1), which will also demonstrate the existence of such decomposition. First pre-multiplying both sides of (2.1) by  $X^T$ , post-multiplying by  $V$  and noting the imposed condition of orthonormality on column vectors of  $U$  and  $V$ , (i.e.  $U^T U = I$  and  $V^T V = I$ , where  $I$  is a unit matrix) we get,

$$RV = X^T X V = V \Gamma U^T U \Gamma V^T V = V \Lambda \tag{2.2}$$

where  $R = X^T X$  is a square symmetric matrix of order  $(p \times p)$  and  $\Lambda = \Gamma^2$  is a diagonal matrix of order  $(r \times r)$  with diagonal elements  $\lambda_i = \gamma_i^2$ . From (2.2), for each diagonal element of  $\Lambda$  we have,

$$(R - \lambda_i I) V_i = 0 \tag{2.3}$$

which could be recognized as a matrix eigenvalue problem. Matrix  $R$  being a cross product of matrix  $X$  of rank  $r$  and being symmetric has  $r$  ( $r \leq \text{Min}\{N, p\}$ ) positive non-zero eigenvalues. Now

the eigenvalues of  $R$  could be computed along with their eigenvectors using standard techniques. The eigenvalues arranged in the descending order of magnitude down the principal diagonal form the matrix  $\Lambda$  and the corresponding normalized eigenvectors arranged from left toward right form the columns of  $V$ . As the columns of  $V$  are the normalized eigenvectors, the orthonormality condition initially imposed on the columns of  $V$  is readily satisfied. Then the matrix  $U$  could be obtained by post multiplying (2.1) by  $V\Gamma^{-1}$ , where  $\Gamma$  is a diagonal matrix with  $\sqrt{\lambda_i}$  along the principal diagonal. This implies that the columns of  $U$  are linear functions of the columns of the data matrix  $X$ . Similarly we can show that the columns of  $V$  can be obtained as linear functions of the rows of  $X$ . Further as imposed initially, the column vectors of  $U$  will be orthonormal. We now show that the column vectors of  $U$  are the eigenvectors of the matrix  $Q = XX^T$ . For this we post multiply both sides of (2.1) by  $X^T U$  and get,

$$XX^T U = U\Gamma V^T V\Gamma^T U^T U = U\Lambda \quad (2.4)$$

where the diagonal elements  $\lambda_i$  of  $\Lambda$  are the eigenvalues of  $R$ . From (2.4), since  $Q = XX^T$ , we can see that all the eigenvalues  $\lambda_i$  of  $R$  and the corresponding column vectors  $U_i$  satisfy the relation,

$$(Q - \lambda_i I)U_i = 0 \quad (2.5)$$

which implies that  $\lambda_i$  and  $U_i$  are respectively the eigenvalues and eigenvectors of the matrix  $Q$ .

We shall discuss here certain useful results connected with the singular value decomposition of a data matrix.

- It follows from the above discussions that  $R$  and  $Q$ , being the cross products of the same matrix  $X$  (i.e.  $R = X^T X$  and  $Q = XX^T$ ), are both symmetric, have the same rank and the same non-zero positive eigenvalues.
- By recalling one of the results of matrix theory, that the sum of all eigenvalues of a square symmetric matrix is equal to the trace of the same matrix, and since the trace of the cross correlation matrix is the sum of squares of all the elements (i.e. the variance) of the matrix  $X$ , we have,

$$\sum_{i=1}^r \lambda_i = Tr\{R\} = Tr\{Q\} = \sum_{i=1}^N \sum_{j=1}^P x_{ij}^2 = Var\{X\} \quad (2.6)$$

where  $Tr\{\cdot\}$  stands for the trace defined as the sum of the elements of the principal diagonal of a square matrix, and the  $Var\{\cdot\}$  stands for the variance of a matrix.

- We can represent the matrix  $X$  as a sum of a series of products of vectors  $U_i$  and  $V_i$  in the form,

$$X = U\Gamma V^T = \sum_{i=1}^r \lambda_i U_i V_i^T \quad (2.7)$$

and the correlation matrices  $R$  and  $Q$  in the form

$$R = X^T X = V \Lambda V^T = \sum_{i=1}^r \lambda_i V_i V_i^T \quad (2.8)$$

$$Q = X X^T = U \Lambda U^T = \sum_{i=1}^r \lambda_i U_i U_i^T \quad (2.9)$$

- Finally the ‘ $r$ ’ orthonormal column vectors of  $V$  can be considered to define an  $r$ -dimensional subspace in the  $p$ -dimensional variable space. Similarly the ‘ $r$ ’ orthonormal column vectors of  $U$  define an  $r$ -dimensional subspace in the  $N$ -dimensional sample space.

## 2.2 Principal Component Analysis

The Principal Component Analysis (PCA) is concerned with the representation of a data matrix in terms of new variables, which are linear combinations of the original variable but are themselves uncorrelated or orthogonal. Let  $X$  be the given data matrix of size  $(N, p)$  and of rank  $r$ . Then, the required new variables can be constructed as linear combinations of the original variables, using the singular value decomposition of the given data matrix. Post multiplying (2.1) by one of the eigenvectors  $V_i$  of the cross correlation matrix  $R = X^T X$ , we get,

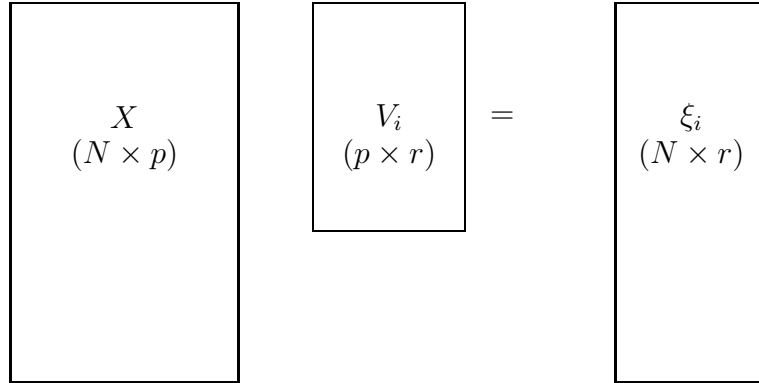
$$X V_i = U \Gamma V^T V_i = U_i \gamma_i \equiv \xi_i \quad (2.10)$$

Evidently the vector  $\xi_i$  is a linear function of the column vectors of the given data matrix  $X$ , and  $\xi_i$  is orthogonal since  $\xi_i^T \xi_j = \gamma_i U_i^T U_j \gamma_j = \gamma_i \delta_{ij} \lambda_j$ , where  $\delta_{ij} = 1$  if  $i = j$  and  $\delta_{ij} = 0$  if  $i \neq j$ . Thus the new vector  $\xi_i$  constructed using an eigenvector  $V_i$  of  $R$ , is the desired principal component. Similarly by post multiplying the matrix  $X$  by different eigenvectors of the matrix  $R$ , we can construct all the columns of the principal component matrix  $\xi$  and arrange them according to the descending order of magnitude of their eigenvalues. Further the column vectors of  $\xi$  are uncorrelated independent vectors. The matrix multiplication defined by (2.10) is illustrated in Figure 2.2.

### 2.2.1 Stationary Vector

The principal components have another useful property. Suppose we look for a set of new vectors as linear functions of the columns of  $X$  and also have stationary values for their variances. Let one of the new vectors be,

**Figure 2.2** Principal Component Analysis



$$\eta = Xa \tag{2.11}$$

where ‘ $a$ ’ is a column vector of size  $p$ . The variance of  $\eta$  is given by,

$$Var\{\eta\} = \eta^T \eta = a^T X^T X a \tag{2.12}$$

It is now required to find ‘ $a$ ’ such that the variance of  $\eta$  is stationary, subject to the condition  $a^T a = 1$ . This problem is equivalent to determining an unconditional stationary value of,

$$L = a^T X^T X a - \lambda (a^T a - 1) \tag{2.13}$$

where  $\lambda$  is the Lagrangian multiplier, an unknown to be determined. For  $L$  to be stationary, we require,

$$(R - \lambda I) a = 0 \tag{2.14}$$

where  $R = X^T X$ . This can readily be recognized as a matrix eigenvalue problem and ‘ $a$ ’ is an eigenvector of the matrix  $R$ . That is,  $\eta$  is the same principal component  $\xi$  defined in (2.10). Note that the above discussions imply that the principal components  $\xi$  are linear function of the column vectors of the given data matrix  $X$ , uncorrelated or orthogonal among themselves and have the property of stationary variance.

Let us express here the data matrix in the form of a series in principal components and also as products of  $U_i$  and  $V_i$  (as in (2.7)),

$$X = \xi V^T = \sum_{i=1}^r \xi_i V_i^T = \sum_{i=1}^r \gamma_i U_i V_i^T \tag{2.15}$$

If we approximate  $X$  by considering only the first  $k$  ( $k < r$ ) terms of the above series, then,

$$X \approx \hat{X} = \sum_{i=1}^k \xi_i V_i^T = \sum_{i=1}^k \gamma_i U_i V_i^T \quad (2.16)$$

and from (2.6),

$$\text{Var}\{\hat{X}\} = \text{Tr}\{\hat{X}\hat{X}^T\} = \text{Tr}\{\xi V^T V \xi^T\} = \text{Tr}\left\{\sum_{i=1}^k \gamma_i^2 U_i U_i^T\right\} = \text{Tr}\{\xi \xi^T\} = \sum_{i=1}^k \lambda_i \quad (2.17)$$

That is the variance of  $\hat{X}$  is equal to the sum of the first  $k$  highest eigenvalues of the matrix  $XX^T$ . Also it follows from the stationary properties of the principal components that  $\text{Var}\{\hat{X}\}$  is the largest possible part of the variance of  $X$  that can be accounted for by taking  $k$  terms in the expansion defined in (2.15). Hence, in the expansion of the matrix  $X$  in terms of  $U_i$  and  $V_i$  in (2.15), the first term,  $\gamma_1 U_1 V_1^T$ , accounts for a maximum percentage of the variance of  $X$ . Similarly the second term in the expansion (2.15) explains the largest percentage of variance of  $X$  left unexplained by the first term and so on. That is, the sum of the first  $k$  products in the expansion (2.15) explains the maximum possible variance of  $X$  and the residual variance left unexplained is minimum. This is also called the least square property of this expansion.

Alternatively by post multiplying  $X^T$  by  $U_i$  (one of the eigenvectors of the matrix  $Q = XX^T$ ), we get another set of principal components of dimension  $p$ ,

$$X^T U_i = V \Gamma U^T U_i = \gamma_i V_i = \zeta_i \quad (2.18)$$

and by post multiplying  $X^T$  by different eigenvectors of the matrix  $Q$ , we can construct all the columns of the principal component matrix  $\zeta$ .

### 2.3 Empirical Orthogonal Function Analysis

Let  $X$  be the given data matrix of size  $(N, p)$  and of rank  $r$ . In Empirical Orthogonal Function (EOF) analysis the data matrix  $X$  is expressed in terms of a set of orthonormal vectors  $U_i$  and  $V_i$ , which are linear functions of the vectors of the matrix  $X$ , in the form,

$$X = U \Gamma V^T \approx \sum_{i=1}^k \lambda_i U_i V_i^T \quad (2.19)$$

where  $k < r$ , the rank of the matrix  $X$ . Since the decomposition of the matrix  $X$  defined in (2.19) is exactly the same as that in SVD analysis, the desired vectors  $U_i$  and  $V_i$  in EOF analysis are determined using the SVD analysis of  $X$ . These vectors are also linearly related to the principal components  $\xi$  and  $\zeta$ . Hence the stationary properties of the principal component expansion of

matrix  $X$  (see (2.15)) prove that the first term  $\gamma_1 U_1 V_1^T$  in (2.19), accounts for the maximum percentage of the variance of  $X$ . Similarly the second term in (2.19) explains the largest percentage of variance of  $X$  left unexplained by the first term and so on. That is EOF generates a minimum number of new variables such that they contain nearly all the essential information in the larger set of observed variables. In this sense EOF is a multivariate method for data reduction or method for removing data redundancies. Generally, this property of EOF often provides a better overview of the underlying salient features of the phenomena observed and may also open up new directions for investigation. EOF analysis is the most frequently used statistical technique in meteorology and oceanography to analyze multivariate data sets of large dimensions describing the space-time distribution of physical variables.

As an example, the repeated measurements of SST values at several stations in an ocean region can be considered to form a data matrix describing a space-time data function/field sampled at limited number of points from the continuously varying field of SST. Consider a data matrix  $X$  of size  $(N, p)$  describing the space-time distribution of physical variable, as in the example cited above. This matrix can be expressed in terms of Empirical Orthogonal Functions/vectors  $U_i$  and  $V_i$  as in (2.19), where the orthonormal vectors  $U_i$  in the  $N$ -dimensional sample space can be considered as a function in time or the temporal EOF. The orthonormal vectors  $V_i$  in the  $p$ -dimensional variable space can be considered as a function describing the spatial distribution or the spatial mode or the spatial EOF of the physical variable of interest. The pattern obtained when a spatial EOF is plotted as a map, represents the mode/ shape of a standing oscillation. The corresponding time function shows how this pattern oscillates in time.

### 2.3.1 Interpretation of EOF

EOF analysis is concerned with the partitioning of the variance of the given data matrix into a series of products of space and time functions. The time functions of the EOF analysis often offer clues to identify the physical mechanism responsible for the variability observed in the data. We can also expect the spatial patterns/ modes of the first few orthogonal functions to be related to possible dynamical modes of physical behavior. But these empirical modes do not necessarily correspond to dynamical modes which conform to physical constraints through governing equations and associated boundary conditions. This is because, often a single physical process may be spread over more than one EOF. In some cases more than one physical process may be contributing to the variance contained in a single EOF. However, EOFs offer the efficient statistical compression of the data field to describe the variability of data sets on a space of reduced dimensions. This is expected to facilitate physical interpretation and whether the empirical modes are related to physical modes is a matter of subjective interpretation.

### 2.3.2 Advantages

- The main feature of EOF is the construction of a few orthogonal vectors in a way that they reduce the overall complexities of the data by taking advantage of inherent inter-dependencies within them. In this sense EOF analysis is a method for removing data redundancies.
- EOF is a tool for describing the variability in the data by a few patterns.
- A minimum number of new variables, which are linear combinations of the original ones, are formed in EOF analysis, such that the new variables contain nearly the same amount of information as the original data set. In this compressed form the new variables, hopefully, bring out the underlying salient features of the phenomena observed.
- A limited number of the first few EOFs can be used to reconstruct the data field, and thereby eliminate those scales of variability that are less energetic in their contribution to the data variance.
- The contribution of any pair of eigenvectors  $(U_i, V_i)$  to the sum of the squares of the elements of the data matrix is given by the corresponding eigenvalue  $\lambda_i$ , which is a measure of relative importance of each pair of vectors. The contribution to this sum from the smaller eigenvalues is sometimes considered as noise.
- It should also be noted that the methods used for achieving all this are basically linear.

### 2.3.3 Shortcomings

- The spatial orthogonality of the eigenvectors often imposes undesirable constraint on the analyzed results (maps). The first map or the spatial eigenvector and the corresponding time function are not influenced by this constraint imposed. But the remaining eigenvectors often bear predictable geometric relationships to the first eigenvector. For example, if the first eigenvector is of one sign throughout the domain, the second will generally be bipolar with its zero line passing through the highest value of the first eigenvector. The third one may be either bi-polar (rotated with respect to the other two) or a more complicated pattern with its maximum near zero lines of the other eigenvectors.

Partially because of the orthogonality constraint, the spatial eigenvectors depend on the particular domain used for the analysis. That is, an eigenmode (physical pattern) over a part of the region may not be the same when the same eigenmode is determined for the whole region. But the invariance of eigenmodes when the domain changes, is a fundamental necessity of the analysis to be physically meaningful. That is, if EOF is not overly sensitive to the area, the physical interpretation of the EOF modes will be more meaningful.

- When individual spatial maps are used to study the correlation between the time series at two grid points, it is possible to have one spatial map indicate that two time series are positively related, while the other map indicate they are negatively correlated. The physical interpretation becomes difficult when the contributions from different spatial maps are significant to the correlation between two time series.
- In EOF analysis, a given function of space and time is always expressed as a sum of a series of products of two functions, one of which depends only on the spatial coordinate and the other function depends only on time (2.19). Such products when considered independently are known to represent standing mode of oscillations. Hence, each product in the series can describe only a standing mode of space-time variation, and propagating waves cannot be detected by this method. This characteristic of EOF severely restricts its usefulness when propagating disturbances are present in the data, as they are expressed as standing waves by EOF analysis.

### Example

Consider a moving disturbance  $h(x, t) = F(\kappa x - \omega t)$ , given in the form of a discrete data set  $H_{ij}$ . Using EOF analysis,  $H_{ij}$  can be expanded in the form,

$$h(x, t) \equiv H_{ij} = T_1(t)X_1(x) + T_2(t)X_2(x) + \dots$$

Each term in the above expansion is a product of two functions, one of which depends only on time and the other only on spatial coordinate. Hence, each product represents the oscillations of a standing mode. Also as each product is examined independently in EOF analysis, the propagating characteristics of the data cannot be detected by this analysis. To be specific consider the example of a wave moving in x direction,

$$h(x, t) = \text{Cos}(\kappa x - \omega t) = \text{Cos}(kx)\text{Cos}(t) + \text{Sin}(\kappa x)\text{Sin}(\omega t)$$

which is also expressed as a sum of products of pairs of functions of space and time. This function as a whole describes a moving wave. But each pair in the above expansion separately describes a standing wave. Through EOF analysis of the discrete form  $H_{ij}$  of this function can be expressed as a sum of two products, each containing a spatial vector and a time vector. The pair of vectors in one of these products will correspond to the pair of functions  $\text{Cos}(\kappa x)$  and  $\text{Cos}(\omega t)$ , evidently describing a standing wave. Similarly the other pair of vector will correspond to  $\text{Sin}(\kappa x)$  and  $\text{Sin}(\omega t)$  and describe the other standing wave. Hence through this analysis, it appears that there are two standing modes instead of a propagating wave. This artifact is introduced purely by the EOF analysis.



The advantages and the shortcomings discussed above are true also for the PCA of multivariate data discussed in this section. Despite these shortcomings, the EOF analysis is found adequate in many applications.

### 2.3.4 EOF Analysis and Statistical Forecasting

EOF analysis has found a central place in statistical forecasting, which essentially attempts to express the future state objectively as a function of its past or present state, without explicit consideration of the nature of the governing dynamics. The orthonormality of the vectors derived through EOF analysis and a fewer number of them required to represent a large fraction of the sum of squares of the elements of a data matrix are desirable features for the construction of statistical prediction equations. Usually the predictand or dependent variable is related to a linear combination of predictor or independent variables using multiple linear regression technique. The predictors may be the past or present values of the predictand or of any other related parameter at the same or different locations.

Let, the predictand be  $y$  and its estimate be  $\hat{y}$ . Then the linear predictor in terms of the  $M$  data parameters  $d_n$  with zero mean will take the form

$$\hat{y} = \sum_{n=1}^M \alpha_n d_n \quad (2.20)$$

where  $\alpha_n$  are the regression coefficients. Defining the optimal predictor/regression formula as the one having the mean square error between  $y$  and  $\hat{y}$  minimum, we have,

$$J = \langle (y - \hat{y})^2 \rangle = \langle y^2 \rangle + \sum_{n=1}^M \sum_{m=1}^M \alpha_n \alpha_m \langle d_n d_m \rangle - 2 \sum_{m=1}^M \alpha_m \langle d_m y \rangle \quad (2.21)$$

where the angle brackets  $\langle \cdot \rangle$  denote the population mean or the expected value, and  $\alpha_n$  can be determined by minimizing  $J$  (i.e. by the method of least squares).

A measure used to determine the closeness with which the predictor estimates the predictand is known as the skill of prediction, and is defined with the help of total and explained variance. The total variance or the sum of the squares of deviations of  $y$  from its mean may be expressed as,

$$\langle (y - \bar{y})^2 \rangle = \langle (y - \hat{y})^2 \rangle + \langle (\hat{y} - \bar{y})^2 \rangle + 2 \langle (y - \hat{y})(\hat{y} - \bar{y}) \rangle \quad (2.22)$$

The first term on the right hand side of the above equation, is the unexplained variance and the second term is the variance of the predicted value or the explained variance, i.e. the variance explained by the regression equation. The last term can be easily shown to be identically equal to zero. Then the skill of prediction with respect to the *population* is given by,

$$S = (\text{Explained variance} / \text{Total Variance}) = \langle (\hat{y} - \bar{y})^2 \rangle / \langle (y - \bar{y})^2 \rangle \quad (2.23)$$

which represents the fraction of  $\langle (y - \bar{y})^2 \rangle$  which is predictable.

Davis [1976] has defined the skill of prediction based on the data sample used for determining the regression coefficients as hind-cast skill ( $S_H$ ), and  $(S_H - S)$  as the artificial predictability. He has proved that  $S_H$  is always greater than  $S$  and the artificial predictability is proportional to the ratio between the number of data parameters  $M$  and the sample size  $N$ .

To reduce the artificial predictability, two avenues are open. One is to increase the sample size, but this is not often possible. The other method is to reduce the number of data parameters/predictors. Lorenz [1956] pointed out that the number of predictors  $M$ , as a thumb rule may be chosen such that  $M/N < 0.1$  to reduce the artificial predictability. Predictors, which are poorly connected with the predictand, may be excluded subjectively by a process of screening. For example, in ocean-atmospheric forecasting problems the predictors, which are not dynamically related to the predictands, could be excluded. An objective method for reducing the number of predictors to a few new independent predictors is EOF analysis.

### 2.3.5 Statistical and Dynamic methods of prediction

Dynamic method of prediction attempts to determine the future state as a complicated nonlinear function of the present state and of the intervening external influences through the solution of governing equations and boundary/ initial conditions. When simplified linearized dynamic equations are used, the fields of interest in the very near future will be a linear function of the present state. Thus the dynamical equations themselves justify the statistical prediction methods. Also it should be mentioned that the dynamic methods as practiced are not entirely free of empirical relations and the approximations made are suggested by the observed behavior of the phenomenon (i.e. available statistics) rather than by pure dynamic theory. Hence we may summarize here that these two methods are not entirely independent of each other, as it appears.

### 2.3.6 Variations in conventional EOF

EOF analysis essentially consists of determining the eigen values (real non-negative) and eigen vectors of the cross-product (covariance/ correlation) matrix, which has real value elements  $a_j$ , and is square and symmetric (i.e.,  $a_{ji} = a_{ij}$ ). An extension of EOF analysis is Complex EOF analysis, which essentially consists of determining the eigen values (real non-negative) and complex eigen vectors of a Hermitian matrix. A Hermitian matrix has complex value elements  $a_j$ , and is square and anti symmetric (i.e.,  $a_{ji} = a_{ij}^*$ , where ‘\*’ indicates the complex conjugate).

While Complex EOF analysis can be used to study the characteristics of propagating waves as well as standing waves, EOF analysis is useful for studying standing waves only. Before the advent of Complex EOF analysis, Cross-spectrum analysis (more details of which are given in

Chapter 3.2) was the standard method for studying geophysical space-time data displaying propagating features.

There are two types of Complex EOF analysis, viz., Complex EOF analysis in frequency domain and Complex EOF analysis in time domain.

In Complex EOF analysis in frequency domain, at a specific frequency band, the eigen values and eigen vectors of the spatial cross-spectral matrix, which is Hermitian are determined. Thus in this analysis (more details of which are given in Chapter 3.4) the first step is Cross-spectrum analysis. Two oceanographic examples where Complex EOF analysis in frequency domain has been used are that of Wang and Mooers [1977] and Michaelson [1982]. This technique was found suitable for analyzing time series in which the energy contained in the data is confined to a few frequency ranges.

In Complex EOF analysis in time domain, the data field to be analyzed is augmented in a manner that helps in detecting the propagating features within it, for which a complex field is constructed from the observed space-time data field. The original data field defines the real part of this complex field, and the imaginary part of this field is defined by the Hilbert transform of the data field. The complex eigenvectors are then determined from the Hermitian cross-product (covariance/ correlation) matrix derived from the complex time series. It may be mentioned that the Hilbert transform of a series has its phase advanced by  $90^\circ$ , i.e., the original time series and its Hilbert transform are in quadrature; however the amplitude remains unaltered. Thus in this analysis (more details of which are given in Chapter 3.5) the first step is the computation of the Hilbert transform of each time series. An excellent review of Complex EOF analysis in time domain is given in Horel [1984]. While the earliest oceanographic application Complex EOF analysis in time domain is that of Barnett [1983], more recent applications are that of White [2000a] and White [2000b]. This technique appears to be more appropriate for data sets with energy spread over a wide range of frequencies.

It is generally accepted, for example by Horel [1984], that Wallace and Dickinson [1972] and Wallace [1972] are the pioneers of Complex EOF analysis in frequency domain. However in our opinion what Wallace [1972] has actually performed appears to be Complex EOF analysis in time domain.

Barnett [1984] has extended Complex EOF analysis in time domain to a vector field - the wind vector. This analysis can be used in case of ocean current vector also. Kundu and Allen [1976] had applied the algorithm of Complex EOF analysis in time domain to the current vector field off Oregon by taking the u-component of the current as the real part and the v-component of the current as the imaginary part. However the approach of Kundu and Allen [1976] appears to be questionable as the u- and v-components of the current are generally not in quadrature [Horel, 1984].

In all the techniques discussed in Chapter 2 and as well as in Chapter 3 for the analysis of complex data matrices, the cross-product matrices are square. But, while analyzing the interdependence between two sets/groups of variables, which are physically related and measured at same number of time levels, the covariance matrix between the two data sets need not be a square matrix. In this case again we use the EOF analysis of covariance matrix to determine the interrelationship between the data fields. The details of this analysis has been discussed under SVD analysis of coupled field in Chapter 4

## 2.4 Factor Analysis

In factor analysis we construct two matrices called factor score and factor loading matrices, each having only a few columns or rows, such that their product contains all the essential information contained in the larger data matrix. The possibility of the existence of such a simple representation is assumed apriori in this analysis. The product of the factor score ( $F$ ) and factor loading ( $A$ ) matrices can be considered as an estimate of the data matrix  $X$ , i.e.,

$$X \approx FA \quad (2.24)$$

and the residual matrix defined as the difference between  $X$  and  $FA$  is small. It should be noted that the aim of the factor analysis is not the exact representation of the original data matrix.

If for example, the data matrix  $X$  is of size  $(N, p)$  and rank  $r$ , then the mathematical representation of R-mode factor model is,

$$X = F_1 A_1 + E_1 \quad (2.25)$$

where  $F_1$  called the factor score matrix is of size  $(N, k)$  and is assumed to have its column vectors (called the factors) orthonormal. Matrix  $A_1$  called the factor loading matrix, is of size  $(k, p)$ .  $E_1$  called the matrix of residual or errors is of size  $(N, p)$ , whose column vectors are orthogonal to those of  $F_1$ .

There are various procedures for the construction of these factor matrices depending upon the additional conditions imposed on these matrices. The frequently imposed condition is that the variance of the matrix  $F_1 A_1$  is maximum. The factor matrices satisfying this constraint can be easily constructed using the principal components  $\zeta$ . Once the number of factors ( say  $k$  factors,  $k < r$ ) is specified, then the first  $k$  columns of the component matrix  $\zeta$  (2.18) can be considered to form the  $k$  rows of the factor loading matrix  $A_1$  and the first  $k$  columns of  $U$  to form the columns of the factor score matrix  $F_1$ . The remaining  $(r - k)$  columns of  $U$  and  $\zeta$  multiply to form the residual matrix  $E_1$ . The above three matrices satisfy perfectly the definition of the R-mode factor model and the additional condition imposed on the variance of  $F_1 A_1$ . As the magnitude of the eigenvalues

generally reduce rapidly from its highest value, only a few factors are needed for sufficiently close representation of the data matrix.

The alternative factor model called the Q-mode factor model is represented mathematically in the form

$$X = A_2 F_2 + E_2 \quad (2.26)$$

where  $F_2$  is the factor score matrix of size  $(k, p)$  and its rows are orthonormal.  $A_2$  is the matrix of factor loading of size  $(N, k)$ . The rows of the residual matrix  $E_2$  are orthogonal to those of  $F_2$ . Let us now impose the condition that the variance of the matrix  $A_2 F_2$  is maximum. Then the factor matrices satisfying this constraint can be easily constructed using the principal components  $\xi$ . The first  $k$  columns of the principal component matrix  $\xi$  (2.10) are chosen to form the columns of  $A_2$  and the first  $k$  columns of  $V$  are chosen to form the rows of the factor score matrix  $F_2$ . The remaining  $(r - k)$  columns of  $\xi$  and  $V$  multiply to form the residual matrix  $E_2$ . Now the matrices  $A_2$  and  $F_2$  form the Q-mode representation of the data matrix.

## 2.5 Varimax Rotation

Consider  $k$  numbers of eigenvectors corresponding to the first ' $k$ ' largest eigenvalues of the cross-product matrix  $(X^T X)$  (where  $k \leq r$ , the rank of the matrix  $X$ ) forming the  $k$  columns of the matrix  $V$ , and the corresponding principal components forming the matrix  $\xi$ . Note that in Q-mode factor analysis,  $\xi$  and  $V^T$  are denoted as factor loading matrix  $A_2$  and factor score matrix  $F_2$  respectively. Then  $V$  and  $\xi$  (i.e.  $F_2^T$  and  $A_2$ ) have the following characteristics (see Chapter 2.2):

- $V^T V = I$ ,
- $\xi_i \equiv X V_i = \gamma_i U_i$ ,
- $\xi^T \xi = \Lambda$  is a diagonal matrix, i.e. column vectors of  $\xi$  are orthogonal among themselves.  $\xi^T \xi$  is also stationary.
- The data matrix  $X$  can be approximated as,  $X \approx \xi V^T$ ,
- $XX^T - \xi \xi^T = \epsilon$ ,
- The diagonal elements of the matrix  $\xi \xi^T$  (i.e.  $\sum_{j=1}^k \xi_{ij}^2$ ) are known as the communality. It is also the variance of the matrix  $X$  explained by each row of the principal component matrix  $\xi$  of  $k$  columns. The diagonal elements of  $\epsilon$  define the unexplained residual variance of  $X$ .

As the principal components are primarily derived to reduce the dimensionality of the data space, the elements of the principal components are often arbitrarily distributed, and the corresponding spatial patterns are difficult to interpret.

The initial principal component solution is then linearly transformed/ rotated according to a suitable criterion to obtain the desired end solution which is easier for interpretation. There are two classes of transformation, orthogonal and oblique, used for this purpose. The orthogonal transformation rigidly rotates a predetermined number of principal components through certain angle in the reduced space determined by the PC/ EOF analysis, for better interpretation of the data. In this method the orthogonality of the reference axes are retained. In oblique rotation method, the principal component vectors are rotated without retaining the orthogonality of the reference axes, to identify the data clusters (if they exist) and to provide greater interpretative capability than orthogonally transformed solutions.

The important difference among the principal components, orthogonally rotated and various obliquely rotated solutions is that the final vectors are described in terms of different sets of reference axes, but in the same Euclidean space of reduced dimensions.

The most frequently used varimax method, which uses the orthogonal rotation of principal components is discussed here in some detail. In this method the variance of the squared rotated components are maximized, unlike in the determination of principal components (where the sum of the squares of the elements of the principal components determined are stationary). The varimax rotation causes the elements / loadings of the rotated principal components to be redistributed, such that a few elements have large amplitudes and many are close to zero, which increases the discrimination among the elements. Hence, in large data matrices, regional relationships between variables (if exists) are enhanced in contrast to the marginal ones over the remaining regions of the domain. This hopefully makes the rotated solutions easier to interpret.

In varimax method the principal components (or factor loading vectors) are linearly combined (or rotated) using *orthogonal* matrices. Let  $W$  be an orthogonal matrix with the property,

$$W^T W = W W^T = I$$

Then the principal components are rotated by  $W$  and the matrix  $X$  is expressed in the form,

$$X \approx \xi V^T = \xi W W^T V^T = \eta \hat{V}^T$$

where  $\hat{V} = V W$  and  $\eta = \xi W$ . As infinitely many orthogonal matrices are possible, it should be noted that the rotated solution  $\eta$  is only one among infinitely many alternative solutions. Then the rotated matrices  $\eta$  and  $\hat{V}$  have the properties:

- $\eta \eta^T = \xi W W^T \xi^T = \xi \xi^T$  i.e. the communality ( $\sum_{j=1}^k \xi_{ij}^2$ ) is invariant under orthogonal rotation

- $\eta^T \eta = W^T \xi^T \xi W = W^T \Lambda W \neq \xi^T \xi$  i.e. rotated column vectors of  $\eta$  are not orthogonal among themselves and lose the principal component property
- and  $\hat{V}^T \hat{V} = I$ .

For the principal components with real elements, the criterion used is that the variance,

$$\sigma^2 = \sum_{j=1}^k \left\{ \sum_{i=1}^N \left[ (\eta_{ij}^2 / h_j^2) - \sum_{i=1}^N (\eta_{ij}^2 / h_j^2) / N \right]^2 \right\} / N$$

is maximum, where  $h_j^2$  is a normalizing factor and represents the variance of the  $j$ -th row of  $k$ -principal components or the communality defined as,

$$h_j^2 = \sum_{n=1}^k \xi_{jn}^2 = \sum_{n=1}^k \eta_{jn}^2$$

where  $k$  is the number of principal components considered in the varimax rotation.

The transformation matrix is determined for two principal components at a time, and then the rotated principal components are obtained. For example, consider the orthogonal rotation of the following two normalized principal components  $x_i$  and  $y_i$  (of row normalized principal component matrix with the square root of their communality),

$$\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_N & y_N \end{bmatrix} \begin{bmatrix} \text{Cos}\phi & -\text{Sin}\phi \\ \text{Sin}\phi & \text{Cos}\phi \end{bmatrix} = \begin{bmatrix} X_1 & Y_1 \\ X_2 & Y_2 \\ \vdots & \vdots \\ X_N & Y_N \end{bmatrix}$$

where  $\phi$  is the angle of rotation and,

$$\begin{aligned} X_i &= x_i \text{Cos}\phi + y_i \text{Sin}\phi \\ Y_i &= -x_i \text{Sin}\phi + y_i \text{Cos}\phi \end{aligned}$$

Now the varimax criterion requires that the sum of the variances of  $X_i$  and  $Y_i$ ,

$$\text{Variance} = \left\{ N \sum_{i=1}^N \left[ (X_i^2) + (Y_i^2) \right] - \left( \sum_{i=1}^N X_i^2 \right)^2 - \left( \sum_{i=1}^N Y_i^2 \right)^2 \right\} / N^2$$

is a maximum. Differentiating the above expression with respect to  $\phi$ , noting that,

$$\begin{aligned} (dX_i/d\phi) &= Y_i \\ (dY_i/d\phi) &= -X_i \end{aligned}$$

and setting the derivative of the *Variance* to zero,

$$N \sum_{i=1}^N [X_i Y_i (X_i^2 - Y_i^2)] - \sum_{i=1}^N (X_i Y_i) \sum_{i=1}^N (X_i^2 - Y_i^2) = 0$$

and substituting for  $X_i$  and  $Y_i$  in terms of  $x_i$  and  $y_i$  and  $\phi$ , we get the angle of rotation  $\phi$  satisfying the varimax criterion as,

$$\phi = (1/4) \text{Tan}^{-1}(\alpha)$$

where  $\alpha = (\alpha_1/\alpha_2)$ , and

$$\begin{aligned} \alpha_1 &= 2 \left[ N \sum_{i=1}^N 2(x_i y_i)(x_i^2 - y_i^2) - \sum_{i=1}^N 2(x_i y_i) \sum_{i=1}^N (x_i^2 - y_i^2) \right] \\ \alpha_2 &= N \left\{ \sum_{i=1}^N [(x_i^2 - y_i^2)^2 - (2x_i y_i)^2] \right\} - \left\{ \left[ \sum_{i=1}^N (x_i^2 - y_i^2) \right]^2 - \left[ \sum_{i=1}^N (2x_i y_i) \right]^2 \right\} \end{aligned}$$

Then both the principal components are rotated through the angle  $\phi$ . This rotation operation is made for all pairs (i.e. for  $k(k-1)/2$  pairs) of principal components in the following order:

$$\begin{aligned} &[(\xi_1, \xi_2), (\xi_1, \xi_3), (\xi_1, \xi_4), \dots (\xi_1, \xi_k); \\ &(\xi_2, \xi_3), (\xi_2, \xi_4), \dots (\xi_2, \xi_k); \\ &\dots \\ &\dots; \\ &\dots (\xi_{k-1}, \xi_k)] \end{aligned}$$

After each rotation, the old components are replaced by the rotated component before the next pair is taken for rotation in the scheme indicated above. The above rotation scheme is repeated until  $k(k-1)/2$  successive rotation of  $\phi = 0$  are obtained, i.e. until the process converges. When the rotation of all the pairs of components is completed the rows of the rotated components are re-normalized by the square root of their respective communalities. The effectiveness of the rotated vectors in identifying the distinct regional features within the data set and helping in their physical interpretation, can be judged by the ratio between the variance ( $\sigma^2$ ) of the rotated principal components and its limiting value, i.e. by the factor,

$$v = 100\sigma^2 k / (k - 1) \tag{2.27}$$

where the limiting value of  $\sigma^2$  has been shown to be  $(k-1)/k$  by Kaiser [1958]. From limited number of studies Horel [1984] reports that if  $v$  is large ( $> 60$ ) the transformation efficiently describes regional factors which are nearly orthogonal, and if  $v$  is small ( $< 40$ ) the varimax approach is inappropriate. It is in order here to remark that, even after rotation of the principal components,



there is no guarantee that the rotated solution will help in the physical interpretation of the solution or represent dynamically important variability. As with any statistical tool, one must determine whether the rotated solutions have any physical meaning.

### 2.5.1 Rotation of Principal components with complex elements

For a complex data matrix, the principal component matrix will be complex (i.e.  $\eta_j$  are complex), and the criterion used in the varimax method is,

$$\sigma^2 = \sum_{j=1}^k \left\{ \sum_{i=1}^N \left[ (\eta_{ij}^* \eta_{ij} / h_j^* h_j) - \sum_{i=1}^N (\eta_{ij}^* \eta_{ij} / h_j^* h_j) / N \right]^2 \right\} / N$$

is maximum, where  $(h_j^* h_j)$  is the normalizing factor and represents the variance of the  $j$ -th row of  $k$ -principal components or the communality.

## 2.6 Number of Components to retain

The number of principal components (Factors/EOFs/modes) retained for rotation can significantly affect the rotated spatial pattern. Hence care should be exercised in selecting a proper number of components (i.e. truncating the eigenvalue sequence) for easy interpretation of the rotated solutions. There are several tests/methods to determine the number of components to be retained for rotation. Horel [1984] suggests that more components than that indicated by typical significance tests need to be retained, as this will not have adverse effects on possible interpretation of components, but retaining less number of components leads to distortion of spatial patterns, which is not easily recognizable.

The variance of principal components has the tendency to decay exponentially for less important components, and hence the curve  $\ln(\lambda_j)$  against the eigenvalue number  $j$  becomes linear at higher modes. Therefore, the eigenvalue sequence can be truncated at the value of  $k$  where the curve becomes linear - this is the simplest method for selecting the principal components for interpretation/rotation!

The lower bound criterion of Guttman [1954] and Hakstian et al. [1982] of retaining all of the principal components that contribute more total variance than does the typical normalised time series, i.e., one unit of total variance, has been used by Horel [1984] in complex principal component analysis.

Preisendorfer and Barnett [1977] and Preisendorfer et al. [1981] suggested a Monte Carlo technique for the selection of number of eigenvectors in a PC/EOF analysis for which the desired signal within the data matrix is above the level of noise. Presented below is a brief summary of one of their significance tests as reported in Overland and Preisendorfer [1982].

Let  $\lambda_j$ ,  $j = 1, \dots, p$ , be the nonzero eigenvalues (arranged in the descending order of their magnitude) of the correlation matrix of the data field. We will form their normalized values  $\hat{\lambda}_j = \lambda_j / (\sum_{k=1}^p \lambda_k)$  and compare each of these normalized eigenvalues with that derived from a spatially and temporally uncorrelated random process. A random number generator is used to generate independent sequence of length  $N$  for  $p$  independent Gaussian variables with zero mean and unit variance. The  $p$  eigenvalues of the cross correlation matrix of the above  $p$  independent variables are computed. This experiment is repeated (say) 100 times. Let the set of normalized eigenvalues produced by the  $r$ -th Monte Carlo experiment be,

$$\hat{\lambda}_j^r = \lambda_j^r / (\sum_{k=1}^p \lambda_k^r), \quad j = 1, \dots, p$$

For fixed  $j$ , let us order the  $\hat{\lambda}_j^r$  so that,

$$\hat{\lambda}_j^1 \leq \hat{\lambda}_j^2 \leq \hat{\lambda}_j^3 \leq \dots \leq \hat{\lambda}_j^{100}$$

For selecting the principal components for which the the geophysical signal is greater than the noise level, the significance test used by Overland and Preisendorfer [1982] is in their own words “Terminate the sequence  $\lambda_j$  at  $j = p'$ , where  $p'$  is the largest integer  $m$  such that  $\hat{\lambda}_m$  exceeds  $\hat{\lambda}_m^{95}$ ”. For convenience Overland and Preisendorfer [1982] have tabulated  $\hat{\lambda}_j^{95}$  for  $j = 1, 2, 3, 4, 5$ ;  $N = 20, 60, 100, 200, 1000$ ; and  $p = 9, 36, 64, 100$  Table 2.1. Overland and Preisendorfer [1982] suggest on the basis of an example concerning cyclone frequencies in the Bering sea, that the tabulated values of  $\hat{\lambda}_j^{95}$ , hold not just in case of eigen values computed from the correlation matrix, but also from those computed from the covariance matrix also.

## 2.7 Correlation matrix or covariance matrix?

While discussing whether to use the correlation matrix or the covariance matrix in EOF analysis, Overland and Preisendorfer [1982] suggest that the covariance matrix is appropriate in locating specific regions with high variance relative to the rest of the field, for example in resolving the spatial distribution of sea surface temperature (SST) anomalies as in Davis [1976]. Correlation matrix on the other hand is more appropriate in cyclone climatology studies, where the primary concern is the study of spatial oscillations or variations of primary storm tracks. Overland and Preisendorfer [1982] state

Comparison of results of the EOF analysis applied to the correlation and covariance matrix of the same data set illustrates an important point in choosing between one or the other approach for application to a geophysical problem. Since the sum of the eigenvalues equals the trace of the matrix, the principal components in the covariance

**Table 2.1** Values of  $\hat{\lambda}_j^{95}$  used in the significance test of Overland and Preisendorfer [1982] for selecting the principal components above the noise level, where  $p$  is the number of spatial locations and  $N$  is the number of elements in each time series

p	j	N				
		20	60	100	200	1000
9	1	29.78	29.78	18.33	15.89	13.11
	2	22.00	17.33	15.67	14.33	12.44
	3	17.89	15.00	14.11	13.22	12.11
	4	14.67	13.22	12.78	12.33	11.67
	5	11.56	11.44	11.56	11.56	11.33
36	1	29.78	29.78	18.33	15.89	13.11
	2	22.00	17.33	15.67	14.33	12.44
	3	17.89	15.00	14.11	13.22	12.11
	4	14.67	13.22	12.78	12.33	11.67
	5	11.56	11.44	11.56	11.56	11.33
64	1	12.00	6.50	5.03	3.86	2.47
	2	10.69	5.89	4.61	3.58	2.38
	3	9.50	5.38	4.34	3.44	2.33
	4	8.78	5.08	4.19	3.28	2.27
	5	7.91	4.77	3.91	3.17	2.23
100	1	10.45	5.31	3.98	2.91	1.74
	2	9.29	4.81	3.72	2.75	1.69
	3	8.57	4.55	3.55	2.65	1.66
	4	7.95	4.30	3.39	2.56	1.62
	5	7.39	4.14	3.23	2.47	1.59

approach are affected by the variance of each spatial variable as well as the covariance between the variables. The covariance approach would therefore be particularly useful in locating specific regions with high variance relative to the rest of the field; an example would be in resolving the spatial distribution of sea surface temperature anomalies. In an application of the correlation matrix the sum of the eigenvalues will again equal the trace of the matrix, but the contribution toward the vectorial direction represented by the EOF's is exclusively from the off-diagonal elements. The spatial-pattern-detection property of the correlation approach, as displayed in contour maps of the EOF modes, is advantageous in such applications as the cyclone climatology, in which one is specifically interested in spatial oscillations or variations of primary storm tracks.