Statistical Methods for Signal Detection in Climate

Silvia A. Venegas

Danish Center for Earth System Science (DCESS),
Niels Bohr Institute for Astronomy, Physics and Geophysics,
University of Copenhagen, Denmark.

DCESS Report # 2

January, 2001

Contents

Chapter 1

Introduction

A climatic signal is, in general terms, the result of interactions among physical processes within the atmosphere-ocean-cryosphere system, which operate on a wide range of spatial and temporal scales. The range of processes involved extend over spatial scales of a few meters to thousands of kilometers and temporal scales of hours to millions of years. The interactions within the components of the climate system usually include positive and negative feedbacks that act on different scales. When these feedbacks combine properly and balance each other, they can give rise to irregular but roughly cyclic climate variations. A well-known example of this is El Niño/Southern Oscillation or ENSO phenomenon.

The motivation for exploratory methods of data analysis in climate comes from the need to separate the climate "signals" from the background climate variability or "noise". This decomposition of the data is done with the hope of identifying the physical processes responsible for the generation of the signal. A fundamental characteristic of the statistical methods for signal detection is their ability to represent spatially distributed data in a compressed way such that the physical processes behind the data, or their effects, can be best visualized by the researcher. Signal detection in climate is useful to achieve four main goals in climate research: a) to recognize the patterns of natural climate variability and distinguish them from presumed anthropogenic or other external (for example solar) effects, b) to use the physical mechanisms inferred from the detected signals to construct numerical climate models, c) to validate numerical climate models by comparing the fundamental characteristics of the modelled data with those of the observed data, and d) to use the signals themselves to forecast the behavior of the system in the future. For all these reasons, the detection and description of climate signals represents a problem of increasing interest in the scientific community.

The complicated behavior and the non-linear character of the climate system provide a real challenge to the exploratory data analysis methods. Climate

variations on widely different timescales, for example, may be connected with one another by nonlinear mechanisms. Some episodic phenomena, such as the climatic response to large volcanic eruptions, seem best suited to be examined in the time domain. Others, such as the periodic seasonal changes in surface temperatures, are better suited to be analyzed in the frequency domain. For certain phenomena it is not clear whether an oscillatory or episodic picture is most appropriate. Also, a number of signals, such as ENSO, exhibit a mixture of time-domain or "event" characteristics and frequency-domain or "oscillatory" characteristics. Such quasi-oscillatory signals are characterized by a dominant timescale of variation, and are often combined with frequency modulation and episodic large-amplitude events. The choice of the appropriate method of analysis is of extreme importance when the objective is to search for specific signals in time, space, or time and space, within large multivariate data sets.

In the last couple of decades, new and more sophisticated methods of data analysis have increasingly been developed and applied in the geophysical community. To a large extent, this has been motivated by the increasing abundance of digital data, observed or generated by models, being collected or produced, and by the equally increasing computational power available to process them. This report attempts to provide an overview of some selected statistical techniques that are commonly applied to the problem of spatio-temporal signal detection in climatic data sets, either observational or model-generated. A considerable amount of the information contained in this report, particularly in the EOF analysis section, was compiled from three fundamental books: Emery and Thomson (1998); von Storch and Zwiers (1999) and von Storch and Navarra (1999).

1.1 Signal and Noise

A fundamental characteristic of the climate data is the high dimensions of the variables representing the state of the system at any given time. Due to this feature, it is often advantageous to be able to split the full phase space into two subspaces, i.e, the *signal* and the *noise* subspaces. The definitions of signal and noise, however, are somewhat arbitrary and largely depend on the interest of the researcher. The term signal is not a well-defined expression in the climate context, therefore the choice of what to call signal and noise is not trivial. In climate research, the signal is defined by the interest of the researcher and the noise is everything else unrelated to this object of interest. In most general terms, a signal can be a pattern in space, or in time, or in space and time, which is determined by the system dynamics. Noise, on the other hand, can be physical or instrumental and comprises all those features and details that

are considered irrelevant for the signal. Generally, the signal has longer scales in space and time than the noise. The signal has also fewer degrees of freedom than the noise.

Let's take as an example the heat transported by the ocean. This is a low-frequency and large-scale process. In the context of the oceanic heat transport, the extratropical storms can be considered noise, since the individual storms do not matter. However, the ensemble of the storms, or the *storm track* is of extreme importance since it controls the energy exchange at the interface of the atmosphere and the ocean. Thus, for some oceanographers, the storm track may constitute the signal and the individual storms are considered noise. For a synoptic meteorologist, however, the individual storm is the object of interest, and thus the signal. To understand an individual storm, the meteorologist does not require a detailed knowledge of each cloud within the storm, so the individual clouds can be considered noise in this context.

The following paragraph, extracted from von Storch and Frankignoul (1998), gives an intuitive and interesting approach to the concepts of signal and noise: "The large amounts of data that are usually studied in climate exhibit a complex mixture of *signals* and *noise*. The purpose of statistical analysis is to disentangle this mixture to find the needle (the signal) in the haystack (the noise). The allegory with the needle in the haystack has two sides. First, it is difficult to find the needle in the haystack. Second, after it has been found, it should be easily recognizable as a needle simply by looking at it. To identify a climatic signal, advanced techniques may be required, but after its identification, the signal usually may be described by means of simple techniques such as composites, correlations, etc."

1.2 Methods of Signal Detection

The statistical techniques described in this review belong to a category of analysis called "Exploratory Analysis". The aim of all these techniques is to summarize the dominant characteristics of a field, such as the dominant space and/or time patterns, and discriminate between the signal of interest and the unrelated processes or noise. We may classify the methods discussed here according to their domain of analysis. Methods of *spatial* pattern detection in multivariate (that is, varying in space and time) data sets are treated in chapter 2. These methods attempt to exploit the information available in spatially distributed data sets and involve eigenvalue decompositions. The most traditional technique is the Empirical Orthogonal Function (EOF) analysis. Two different approaches are given for performing the EOF decomposition: the covariance matrix approach and the singular value decomposition approach. Discussions on the physical interpretation of EOFs and on rotation of EOFs

are also given. In addition, three variations on the classical EOF analysis are described that serve to jointly analyze two or more fields. These are the Combined EOF analysis, the Singular Value Decomposition of coupled fields (SVD) and the Canonical Correlation Analysis (CCA).

Time series analysis methods attempt to detect and isolate the *temporal* signals that are blended together inside a single time series (univariate methods), and are described in chapter 3. The Multi Taper Method (MTM) is a technique for spectral analysis that provides an estimate of the power spectrum of a time series with optimal spectral resolution and variance properties, by tapering the time series with a set of orthogonal tapers that are resistant to spectral leakage. The Singular Spectral Analysis (SSA) is a time series analysis technique used to detect recurrent temporal patterns in univariate data series. It is in fact a variation of the classical EOF analysis, in which the decomposition is performed in the time domain. The Wavelet Analysis (WA) makes use of a Wavelet Transform to provide a two-dimensional power spectrum that displays the evolution of the amplitude at different frequencies with time.

Further extensions of the EOF technique are designed to identify the dominant *spatial* and *temporal* structures in a multivariate data set. These methods are well suited to investigate propagating features since they retain phase information in the decompositions. They are discussed in chapter 4. The Extended EOF analysis (EEOF) decomposes a lagged covariance matrix derived from the original time series sampled at different temporal lags. The Frequency Domain EOF analysis (FDEOF) is only briefly outlined here for completeness, since its very general approach is not especially suitable for climate applications. As a derivate of FDEOF, the alternative Complex or Hilbert EOF analysis (CEOF) is described in detail as a simple and effective approach to investigate propagation of signals. The Multichannel SSA (MSSA) technique is also briefly described as the multivariate generalization of the SSA method. In fact, the MSSA procedure is mathematically equivalent to EEOF analysis. Finally, a combination of the Multi Taper Method with the Singular Value Decomposition, called the MTM-SVD technique, is also described. This recently developed multivariate frequency-domain decomposition seeks to isolate statistically significant narrow-band oscillations that are correlated among a number of spatial locations. The MTM-SVD approach allows for the detection and description of oscillatory signals that may be modulated in frequency, phase and amplitude.

Chapter 2

Patterns in Space: Multivariate Analysis

2.1 Empirical Orthogonal Function (EOF) Analysis

It is usual in climate studies to be presented with a large data set consisting of time series over a grid of stations which we wish to compress into a smaller number of independent pieces of information. Typically it is necessary to deal with an ensemble of instantaneous samples (maps) of a geophysical field (for example, temperature) defined at a number of points (stations). In such cases, the data is in the form of simultaneous time series records from a grid on a horizontal plane: $x_i(t), y_i(t)$. The grid points may be regularly spaced (such as model-generated data or gridded observations) or irregularly spaced (such as locations of meteorological/oceanographic stations). The data may also be in the form of time series on a cross-section on a vertical plane: $x_i(t), z_i(t)$. In this case, the data could come from a selection of depths in the ocean (such as current meter measurements), or at standard pressure levels in the atmosphere (such as geopotential height or air temperature measurements). Analyses of data sets with the described characteristics, that is, consisting of a number of spatially distributed time series, are known as multivariate analyses. The method of Empirical Orthogonal Functions (EOF), also known as Principal Component Analysis (PCA) is a particularly useful technique for compressing the variability in this type of data sets. The EOF approach was first introduced in fluid dynamics by Edward Lorenz in 1956, and has been widely applied in meteorology and oceanography studies since then.

The goal of the EOF analysis is to provide a compact description of the spatial and temporal variability of data series in terms of orthogonal functions or statistical "modes". Usually, most of the variance of the spatially distributed

time series is in the first few orthogonal functions whose patterns may then be linked to possible dynamical mechanisms. It should be emphasized that no direct physical relationship necessarily exists between the purely statistical EOFs and any *a posteriori* related dynamical process. EOF analysis is simply a method for partitioning the variance of a spatially distributed group of time series. They are called *empirical* to reflect the fact that they are defined by the covariance structure of the specific data set being analyzed. A more extended discussion on the physical interpretation of the EOFs can be found in section 2.1.5.

There are two approaches for computing EOFs for a number of time series. The first constructs the covariance matrix of the data series and then decomposes it into eigenvalues and eigenvectors. The second uses the singular value decomposition of the data matrix to obtain eigenvalues, eigenvectors and time varying amplitudes (principal components) without computing a covariance matrix. The EOFs obtained from the two methods are identical. The main difference between the two is given by the greater degree of sophistication, computational speed and stability of the singular value decomposition approach. However, since traditionally the covariance matrix method has been largely used, both approaches are discussed in details in the next sections, after a preliminary discussion on the preparation of the data.

2.1.1 Preparation of the Data

Let's consider a set of N maps at times $t = 1 \ldots N$, where each map contains measurements of the field ψ at locations $m = 1 \ldots M$. That is, we have M time series $\psi_m(t)$, each of length N. The first $(t = 1)$ and last $(t = N)$ times should be the same for all the M series. Here we will assume that $N > M$, that is, the number of samples (time steps) is larger than the number of locations (time series) as it is often the case in geophysical datasets. However, the case $N < M$ is also possible and an alternative solution for this case is given in section 2.1.3.

In analyses of geophysical fields, we are often interested in variability on timescales other than the annual, that is, if the data is known to include an annual (seasonal) cycle, its detection is usually irrelevant to our results. In this case, we would like to remove the seasonal variations *before* doing the EOF analysis. This can be done by computing the climatological annual cycle and by subtracting it from the field $\psi_m(t)$. We are thus left with the deviations from the annual cycle, or the *anomalies*. This procedure may be generalized to remove other well-known signals on other timescales which we explicitly want to exclude from the analysis for any particular reason. Therefore, field $\psi_m(t)$ may either be the complete data or the anomalies, depending on the choice of the researcher and the kind of study to be performed.

We first remove the record mean μ_m from each time series. We may optionally also compute and remove any linear trend existing in the record, as long as we are not interested in it. Second, we normalize the demeaned (optionally detrended) time series by dividing each of them by its standard deviation σ_m. This ensures that the analysis is not dominated by the variance from any given location (all locations are given an equal chance to contribute in the analysis). The resulting demeaned and normalized series $F_m(t)$ are termed "standarized" series:

$$F_m(t) \;=\; \frac{\psi_m(t) \;-\; \mu_m}{\sigma_m} \tag{2.1}$$

where μ_m is the record mean:

$$\mu_m \;=\; \frac{1}{N} \sum_{t=1}^{N} \psi_m(t) \tag{2.2}$$

and σ_m is the record standard deviation:

$$\sigma_m \;=\; \left[\frac{1}{N-1} \sum_{t=1}^{N} \psi_m^2(t) \right]^{1/2} \tag{2.3}$$

The normalization is especially relevant when analyzing two or more fields together, to ensure no dominance of one field over the others (this is important in sections 2.2, 2.3 and 2.4). Several single-field studies, however, have reported that the normalization makes only a negligeable difference in the resulting patterns, especially when the variable analyzed has a relatively uniform distribution of variance over the different spatial locations. Nevertheless, it is recommended to normalize the data to variance one for simplicity when analyzing the results.

After this preliminary manipulation of the time series, the data set is ready to be used as input for the EOF analysis. We construct an $M \times N$ data matrix $\mathbf{F}$ by organizing the M rows (locations m) and the N columns (times t) of the standarized data (or anomalies) as follows:

$$\mathbf{F} = \begin{array}{c} \text{Time} \longrightarrow \\ \left[\begin{array}{cccc} F_1(1) & F_1(2) & \dots & F_1(N) \\ F_2(1) & F_2(2) & \dots & F_2(N) \\ \dots & \dots & \dots & \dots \\ F_M(1) & F_M(2) & \dots & F_M(N) \end{array}\right] \end{array} \begin{array}{l} \\ \\ \downarrow \text{Location} \end{array} \tag{2.4}$$

The treatment of the data matrix $\mathbf{F}$ now depends on the approach taken to perform the EOF decomposition. The following sections deal with the two mentioned approaches: the first computes a covariance matrix and the second uses the singular value decomposition.

2.1.2 The Covariance Matrix Approach

Data matrix $\mathbf{F}$ is now used to derive the spatial covariance matrix $\mathbf{R_{FF}}$ of the field $F_m(t)$ by multiplying matrix $\mathbf{F}$ by its transpose $\mathbf{F}^\dagger$:

$$\mathbf{R_{FF}} = \mathbf{F} * \mathbf{F}^\dagger \tag{2.5}$$

Expanding the product of matrices:

$$\mathbf{R_{FF}} = \left[\begin{array}{cccc} \langle F_1 F_1 \rangle & \langle F_1 F_2 \rangle & \dots & \langle F_1 F_M \rangle \\ \langle F_2 F_1 \rangle & \langle F_2 F_2 \rangle & \dots & \langle F_2 F_M \rangle \\ \dots & \dots & \dots & \dots \\ \langle F_M F_1 \rangle & \langle F_M F_2 \rangle & \dots & \langle F_M F_M \rangle \end{array}\right] \tag{2.6}$$

where $\langle F_i F_j \rangle$ is the covariance between time series F_i and F_j (F at locations i and j) defined as:

$$\langle F_i F_j \rangle \;=\; \langle F_j F_i \rangle \;=\; \frac{1}{N-1} \sum_{t=1}^{N} F_i(t)\, F_j(t) \tag{2.7}$$

where $i, j = 1 \dots M$. The matrix product $\mathbf{R_{FF}}$ is symmetric and square, even if $\mathbf{F}$ itself is not square. The dimension of $\mathbf{R_{FF}}$ is $M \times M$. If the data series in $\mathbf{F}$ are normalized by the standard deviation as suggested above, the matrix $\mathbf{R_{FF}}$ is formally the correlation matrix instead of the covariance matrix.

We want to remark here that some authors define the data matrix $\mathbf{F}$ as the transpose of that we defined in equation 2.4, that is, with M *columns*

corresponding to locations and N *rows* corresponding to time steps. In such a case, the determination of the spatial covariance matrix should be done as $\mathbf{R_{FF}} = \mathbf{F}^{\dagger} * \mathbf{F}$. The rest of the procedure, however, is identical to what is described here. Also, some texts define the covariance by dividing by N instead of $N-1$ in equation (2.7). This will only affect the results in that the EOFs are the same to within a constant factor.

Once the covariance matrix has been calculated from the data, we need to solve the eigenproblem:

$$\mathbf{R_{FF}} * \mathbf{E} = \mathbf{E} * \mathbf{\Lambda} \tag{2.8}$$

That is, we decompose $\mathbf{R_{FF}}$ into matrices $\mathbf{\Lambda}$ and $\mathbf{E}$. Here $\mathbf{\Lambda}$ is the $M \times M$ diagonal matrix containing the eigenvalues λ_k of $\mathbf{R_{FF}}$:

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_M \end{bmatrix} \tag{2.9}$$

The eigenvalues in $\mathbf{\Lambda}$ are usually sorted in decreasing order, so that $\lambda_1 > \lambda_2 > \dots > \lambda_M$. Also, since the data matrix $\mathbf{F}$ is real, the covariance matrix $\mathbf{R_{FF}}$ is positive definite, which means that all its eigenvalues are greater or equal to zero. Although the dimension of matrix $\mathbf{\Lambda}$ is $M \times M$, typically only the first K eigenvalues λ_k, $k = 1 \dots K$ are non-zero, where $K \leq min(N, M)$. Hence, the "effective" dimension of $\mathbf{\Lambda}$ is in fact $K \times K$. This implies that only K EOF modes can be determined. In the following, the index k will represent the "mode".

The square matrix $\mathbf{E}$ has dimension $M \times M$. Its column vectors E^k are the eigenvectors of $\mathbf{R_{FF}}$ corresponding to eigenvalues λ_k:

$$\mathbf{E} = \begin{bmatrix} E_1^1 & E_1^2 & \dots & E_1^M \\ E_2^1 & E_2^2 & \dots & E_2^M \\ \dots & \dots & \dots & \dots \\ E_M^1 & E_M^2 & \dots & E_M^M \end{bmatrix} \tag{2.10}$$

$$\downarrow \quad \downarrow \qquad \downarrow$$

$$E^1 \quad E^2 \qquad E^M \longrightarrow \text{Eigenvectors } E^k$$

Each non-zero eigenvalue λ_k in matrix $\mathbf{\Lambda}$ is associated with a column eigen-

vector E^k in matrix $\mathbf{E}$. Therefore, only K eigenvectors are used in the decomposition, those corresponding to the K non-zero eigenvalues. As such, the effective dimension of matrix $\mathbf{E}$ is $M \times K$, where M are the spatial locations and K are the modes of the EOF decomposition.

The eigenvector matrix $\mathbf{E}$ has the property that $\mathbf{E} * \mathbf{E}^\dagger = \mathbf{E}^\dagger * \mathbf{E} = \mathbf{I}$, where $\mathbf{I}$ is the Identity matrix. This means that the eigenvectors are uncorrelated over space, that is, they are orthogonal to one another. Each eigenvector E^k represents the spatial EOF pattern of mode k (it has dimension M, that is, the number of locations in the original maps). We will refer to the spatial EOF patterns as the *Empirical Orthogonal Functions* or simply EOFs. Other names found in the literature are Principal Vectors or Loadings.

The time evolution of the kth EOF (that is, how pattern E^k evolves with time) is given by the time series $A^k(t)$, which is obtained by projecting the original data series $F_m(t)$ onto eigenvector E^k and summing over all locations m:

$$A^k(t) \;=\; \sum_{m=1}^{M} E_m^k \; F_m(t) \tag{2.11}$$

where $m = 1 \ldots M$ counts the locations, $t = 1 \ldots N$ counts the time steps and $k = 1 \ldots K$ counts the EOF modes. In matrix notation, matrix $\mathbf{A}$ is obtained by multiplying matrices $\mathbf{E}^\dagger$ and $\mathbf{F}$:

$$\mathbf{A} = \mathbf{E}^\dagger * \mathbf{F} \tag{2.12}$$

where $\mathbf{E}^\dagger$ (the transpose of $\mathbf{E}$) is $K \times M$, $\mathbf{F}$ is $M \times N$, and hence $\mathbf{A}$ is $K \times N$. Note that we have used only the "effective" matrix $\mathbf{E}$ to reduce the amount of data in the computations. Rows in matrix $\mathbf{A}$ are time series of length N, that is the number of time steps in the original time series. We will refer to them as the *Principal Components* or PCs. Other common names found in the literature are time series of Expansion Coefficients, Time Coefficients, Eigenvector time series or Scores. Just as the spatial patterns E^k are orthogonal in space, the principal components A^k are orthogonal in time.

Each eigenvalue λ_k is proportional to the percentage of the variance of the field F that is accounted for by mode k. This percentage is calculated as:

$$\%\ \text{Variance Mode k} \;=\; \frac{\lambda_k}{\sum_{i=1}^{K} \lambda_i} * 100 \tag{2.13}$$

The original field F can be totally reconstructed by multiplying each EOF pattern E^k by its corresponding principal component A^k and adding the products over all K modes:

$$F_m(t) \;=\; \sum_{k=1}^{K} E_m^k \; A^k(t) \tag{2.14}$$

In matrix notation:

$$\mathbf{F} = \mathbf{E} * \mathbf{A} \tag{2.15}$$

where $\mathbf{E}$ is $M \times K$, $\mathbf{A}$ is $K \times N$, and hence $\mathbf{F}$ is $M \times N$. However, the goal of the EOF decomposition is in fact the reconstruction of an approximate, compressed and less noisy version $\hat{F}$ of the original field F. This is done by truncating the decomposition in equation (2.14), that is, by reconstructing $\hat{F}$ using only the first H modes, with $H < K$. The H first modes account for the largest fraction of the field variance:

$$\hat{F}_m(t) \;=\; \sum_{k=1}^{H} E_m^k \; A^k(t) \tag{2.16}$$

This leads to a significant reduction of the amount of data while retaining most of the variance of field F. The special characteristics of this truncation and the rather subjective choice of H are further discussed in section 2.1.5. *Sometimes*, the first or the few first EOF modes represent meaningful physical processes, which are associated with characteristic spatial and temporal patterns. However, this is not necessarily so. The physical interpretation of EOFs is a highly relevant topic and it will be extensively discussed in section 2.1.5.

2.1.3 Alternative when $M > N$

In the above we have assumed that $M < N$, that is, the number of locations in the data set is less than the number of time steps (or maps/samples) available. We may have, for example, 100 years of monthly air temperature observations ($N = 1200$) measured at $M = 50$ meteorological stations, which resembles the size of a typical dataset in meteorology and oceanography. However, there are examples of datasets for which the ratio between the M and N dimensions is the inverse. For satellite images, for instance, there may be $M = 5000$ spatial

points around the world sampled $N = 50$ times. In such a case, the spatial covariance matrix as calculated in section 2.1.2 has dimension 5000×5000, which requires extremely large computational power to be decomposed. An alternative solution to this is based on the following algebraic trick: matrix $\mathbf{R_{FF}} = \mathbf{F} * \mathbf{F}^\dagger$ of dimension $M \times M$ and matrix $\mathbf{R^\dagger_{FF}} = \mathbf{F}^\dagger * \mathbf{F}$ of dimension $N \times N$ share the same non-zero eigenvalues. In the case of having $M > N$, matrix $\mathbf{R^\dagger_{FF}}$ is smaller than $\mathbf{R_{FF}}$. Hence, we may just calculate the smallest of the two covariance matrices, $\mathbf{R^\dagger_{FF}}$, and solve the eigenvalue problem on it as in equation (2.8):

$$\mathbf{R^\dagger_{FF}} * \mathbf{D} = \mathbf{D} * \mathbf{\Lambda} \tag{2.17}$$

The eigenvalues in matrix $\mathbf{\Lambda}$ are the same as those obtained from equation 2.8 (that is, matrix $\mathbf{\Lambda}$ is equal to matrix (2.9)). The eigenvectors in matrix $\mathbf{D}$ obtained from the decomposition of the small covariance matrix $\mathbf{R^\dagger_{FF}}$, however, are *not* the same as those in $\mathbf{E}$ obtained from the decomposition of the covariance matrix $\mathbf{R_{FF}}$ (computed in section 2.1.2). Nevertheless, we can obtain the eigenvector matrix $\mathbf{E}$ by projecting matrix $\mathbf{D}$ onto field $\mathbf{F}$ as:

$$\mathbf{E} = \mathbf{F} * \mathbf{D} \tag{2.18}$$

where $\mathbf{F}$ is $M \times N$, $\mathbf{D}$ is $N \times N$ and hence $\mathbf{E}$ is $M \times N$. Here we observe that, using this alternative method, we can only compute the first N EOF modes of the total possible M modes of the M-variate field $\mathbf{F}$ (since matrix $\mathbf{E}$ is $M \times N$ instead of $M \times M$ as in section 2.1.2). However, this is not strictly a problem in our applications, since we are only interested in a few leading modes (two or three at most).

2.1.4 The Singular Value Decomposition Approach

There are two good reasons for choosing the singular value decomposition approach rather than the covariance matrix approach in order to perform EOF analysis. First, it provides a one-step method for computing all the components of the eigenvalue problem, without having to compute and store large covariance matrices. Second, the resulting decomposition is computationally more stable and robust.

In this approach, a singular value decomposition is performed directly on the rectangular data matrix $\mathbf{F}$ (matrix (2.4)), constructed in section 2.1.1, that consists of M rows (spatial points) and N columns (temporal samples). The

singular value decomposition of a matrix is based on the concept that any rectangular $M \times N$ matrix $\mathbf{F}$ can be written as the product of three matrices: an $M \times M$ matrix $\mathbf{U}$, a $M \times N$ diagonal matrix $\mathbf{\Gamma}$ with positive or zero elements, and the transpose ($\mathbf{V}^\dagger$) of the $N \times N$ matrix $\mathbf{V}$.

$$\mathbf{F} = \mathbf{U} * \mathbf{\Gamma} * \mathbf{V}^\dagger \tag{2.19}$$

Matrix $\mathbf{\Gamma}$ is a rectangular $M \times N$ matrix with zero elements outside the diagonal and positive or zero elements on the diagonal. The scalars on the diagonal, γ_k, are called the *singular values* and are typically placed in decreasing order of magnitude. The singular values γ_k are proportional to the eigenvalues λ_k obtained in section 2.1.2 (matrix (2.9)) such that $\lambda_k = {\gamma_k}^2$. Again, there is a maximum of $K \leq min(N, M)$ non-zero singular values, which defines the maximum number of EOF modes we can determine, so that the effective dimension of matrix $\mathbf{\Gamma}$ is $K \times K$.

The columns of the quadratic $M \times M$ matrix $\mathbf{U}$ are orthogonal and are called the *left singular vectors* of $\mathbf{F}$. They are identical to the eigenvectors E (matrix (2.10)) obtained from equation (2.8), that is, they are the EOF patterns associated with each singular value. Again, there are only K useful left singular vectors associated with the K non-zero singular values, hence the effective dimension of matrix $\mathbf{U}$ is $M \times K$.

The rows of the quadratic $N \times N$ matrix $\mathbf{V}^\dagger$ are also orthogonal and are called the *right singular vectors* of $\mathbf{F}$. They are proportional to the principal components A obtained from equations (2.11)-(2.12), and the constant of proportionality are the singular values γ_k such that:

$$\mathbf{A} = \mathbf{\Gamma} * \mathbf{V}^\dagger \tag{2.20}$$

$$A^k(t) = \gamma_k \, V^{\dagger k}(t) \tag{2.21}$$

where the effective size of $\mathbf{A}$ is $K \times N$. Matrix $\mathbf{A}$ contains the principal components of data matrix $\mathbf{F}$. Using equation (2.19) we can reconstruct field F adding all K modes of the decomposition as:

$$F_m(t) = \sum_{k=1}^{K} U_m^k \, \gamma_k \, V^{\dagger k}(t) \tag{2.22}$$

Note the similarity between this equation and equation (2.14) (in the covari-

ance matrix approach), where $U = E$ and $A = \gamma V^{\dagger}$.

As we discussed in section 2.1.2, some authors define the data matrix $\mathbf{F}$ as the transpose of that we defined here, that is, with M columns corresponding to locations and N rows corresponding to time steps. This does not really make a difference in the singular value decomposition approach, except for the fact that, in such a case, the eigenvectors of $\mathbf{F}$ are found in the right matrix $\mathbf{V}^{\dagger}$ and the principal components are proportional to the left matrix $\mathbf{U}$.

2.1.5 Physical Interpretation of EOFs

The relative importance of the EOF modes obtained above may be measured by their capability to account for the total variance of $F_m(t)$. As we said, each EOF mode has an associated eigenvalue. The larger the eigenvalue, the larger the portion of the variance accounted for by the mode. EOF modes are ordered by decreasing eigenvalue so that the first mode, having the largest eigenvalue, accounts for the largest percentage of the variance of the data. Based on the inherent efficiency of this statistical decomposition, a set of very few EOFs is generally enough to describe the fundamental variability within a large dataset. Often it can be advisable to use EOFs as a filter to eliminate undesired scales of variability. In such a case, a limited number of EOFs are used to reconstruct a good approximation of the original dataset. By doing so, we can filter out the scales of variability that are not coherent over the entire data field, and that therefore are less important in their contribution to the field variance.

The question that now arises is: what is an adequate truncation of the EOF decomposition of field $F_m(t)$? This question may also be formulated as: which are the physically significant EOF modes? Unfortunately, the answer to this question is not obvious and depends on the specific problem we are trying to solve. In general, a relevant piece of information is provided by the amount of variance accounted for by each EOF. We might select a subset of H EOF modes ($H < K$) so that their accumulative accounted variance reaches a certain threshold (typical threshold values can be 80% or 90%). Or such that the last kept EOF (the Hth mode) accounts for a certain minimum portion of variance (for example 5% or 1%). A good choice of H requires that the bulk of the variance of field $F_m(t)$ can be represented by the first H EOFs. If the original variable has M components, the approximation of $F_m(t)$ by H EOFs ($H < M$) leads to a significant reduction of the amount of data while retaining most of the variance.

In interpreting the meaning of EOFs, we need to keep in mind that, while EOFs offer the most efficient statistical compression of the data field, they do not necessarily correspond to real dynamical modes or modes of physical behavior. It is often the case that a single physical process is spread over several

EOF modes. In other cases, more than one physical process may contribute to the variance contained in a single EOF. The statistical modes derived from the EOF procedure must be considered in light of accepted physical mechanisms rather than as physical modes themselves. It is often likely that the strong variability associated with the dominant modes is attributable to recognizable physical mechanisms. A clue to the physical mechanisms associated with an EOF mode may often be found in the principal component $A(t)$. Something may be known about the temporal variability of a process that might resemble the principal component. This would then suggest a relationship that is not clearly apparent looking at the spatial EOF structure only. The EOF patterns are constructed to represent in an optimal manner field variance and *not* physical connections or maximum correlations. They are excellent tools to compress data into a few significant components selected according to the variance explained. However, the patterns that most efficiently represent variance do not necessarily have anything to do with the underlying dynamical structure.

The physical interpretation of EOFs is also limited by a fundamental constraint: the imposed spatial orthogonality of the EOF patterns and the resulting temporal independence of the time coefficients. While it is often possible to associate the first EOF mode with a known physical process, this is much more difficult with the second (and higher order) EOF because it is constrained to be orthogonal to the first EOF. The orthogonality constraint implies that the EOF modes can only represent physical processes that operate independently and with orthogonal patterns. Real world processes, however, do not need to have orthogonal patterns or uncorrelated indices. On the contrary, they are in general interrelated.

Traditional EOF analysis can only detect standing oscillations. However, in some special cases, we may anticipate the presence of a propagating signal with a classical EOF decomposition if two principal components $A^k(t)$ and $A^{k+1}(t)$, associated with two consecutive eigenvalues λ_k and λ_{k+1}, vary coherently and are in quadrature, that is, are 90 degrees out of phase with one another. In this case, the eigenvalues have similar magnitude and the pairs of EOFs and PCs may represent a signal that is propagating in space. If such an indication of propagation is given by an EOF analysis, it is advisable to perform a variation of the EOF technique suited to detect propagating features, such as EEOF or CEOF (see sections below), to confirm the result.

The EOF patterns are sometimes dependent on the size of the domain of study. If we deal with a variable with relatively uniform distribution of variance and we know that the *characteristic spatial scale* of the variable is comparable to, or larger than, the considered spatial domain, then the first EOF will in *most* cases be a monopole pattern, that is, a spatial pattern with the same sign at all points. This is simply because the system will have a tendency to create anomalies of the same sign in the entire domain when the typical scale

of the field is as large as the domain of study. The need to be orthogonal to the first EOF then creates a second EOF with a dipole pattern, which is the largest-scale pattern orthogonal to the uniform-sign first EOF. If, however, the characteristic spatial scale of the variable is smaller than the analysis domain, then the first EOF does not necessarily result in a monopole pattern. It is thus advisable that, when possible, the size of the domain of analysis be greater than the characteristic spatial scale of the field to be analyzed.

According to Peixoto and Oort (1992), one way to understand the basic idea behind EOFs is to imagine that we can display each of the N maps as a vector f_n in the M-dimensional space, such that:

$$f_n = \{f_{n1}, f_{n2}, \ldots, f_{nM}\} \quad \text{at time t = n} \tag{2.23}$$

Each vector f_n includes the values of field f at all locations $m = 1, \ldots, M$ for a given time n (that is, a map of f at time n). Each of the N data vectors is directed from the origin to a point in the M space (see Figure 2.1). If there exists some correlation between the N vectors (maps), we expect that their extremities will be organized in clusters or along some preferred directions. The problem we want to solve with the EOF decomposition is to find an orthogonal basis $\{e_1, e_2, \ldots, e_M\}$ in the M-dimensional space, instead of the original basis, such that vector e_1 best represents the largest cluster of the original data vectors, e_2 best represents the second largest cluster of the original data vectors, and so on. In other words, e_1 accounts for the largest portion of the data variance, e_2 for the second largest portion, and so on. This is equivalent to finding a set of M vectors e_m whose orientation is such that the sum of the squares of the projections of all the N observation vectors f_n onto each e_m is maximized sequentially. The vectors e_m, $m = 1, \ldots, M$ are mutually orthogonal and they are what we called the EOFs.

All the above discussions about interpretation of EOFs are also applicable to the EOF variations on the EOF method described in the next sections. Further reading about EOF analysis can be found in Preisendorfer (1988); Emery and Thomson (1998); von Storch (1999) and von Storch and Zwiers (1999). See also the references for each particular EOF variation given in the following sections.

2.1.6 Units and Presentation

Formally, the units of the field F are carried by the principal components while the EOF patterns are dimensionless. However, in practice, it is common to re-normalize the PCs and the EOFs such as:

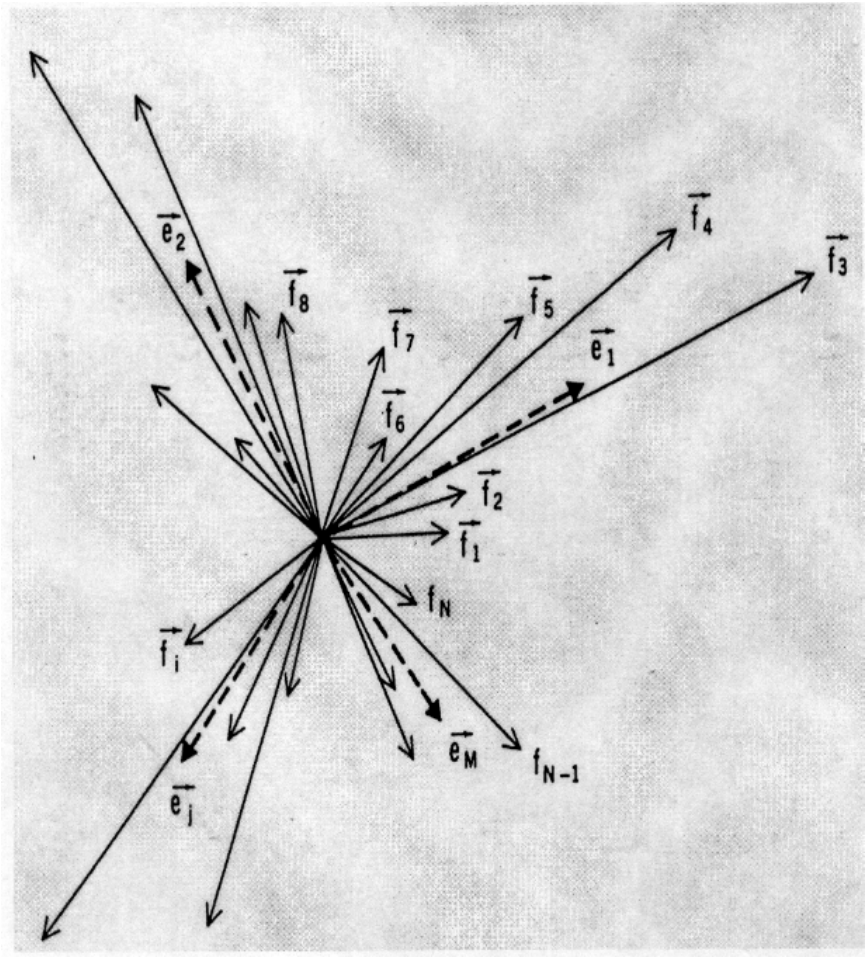

Figure 2.1: Example of a possible configuration of the data vectors f_n ($n = 1 \ldots N$ denote the time steps) and the empirical orthogonal vectors e_m, $m = 1 \ldots M$. From Peixoto and Oort (1992)

$$A^{k*}(t) = \frac{A^k(t)}{\sqrt{\lambda_k}} \tag{2.24}$$

$$E_m^{k*} = E_m^k \sqrt{\lambda_k} \tag{2.25}$$

where λ_k is the kth eigenvalue of the decomposition. The re-normalized EOFs E_m^{k*} then carry the units of F and the PCs $A^{k*}(t)$ have variance 1. As such, the EOFs represent the amplitude of a "typical" event when the PCs are ± 1 (von Storch and Zwiers, 1999).

An alternative to the re-normalization is the presentation of the EOF patterns as correlation maps. A correlation map for mode k is a map of correlation coefficients $\mathbf{r}_m^k$ between the principal component $A^k(t)$ and the values of field $F_m(t)$ at each location $m = 1 \ldots M$:

$$\mathbf{r}_m^k = [A^k(t),\ F_m(t)] \tag{2.26}$$

where [ ] indicates temporal correlation, k indicates the mode and m indicates the location. The contours of such a map show the distribution of the centers of action of the mode scaled as correlation coefficients, which is more meaningful than the dimensionless EOFs in vector E. The distribution of centers of action in the correlation map is basically the same as that in the EOF spatial pattern.

2.1.7 Rotation of EOFs

As we have seen, EOFs are designed to be the most efficient descriptors of the variability in a data set. However, this does not necessarily mean that they lead to the clearest physical interpretation of the processes behind the data set, since different modes of variability in the real world need not be orthogonal in space and time as EOFs are. A "rotation" of the EOFs may help overcome some of these difficulties. The general concept of rotation is to replace the EOF patterns E obtained in equation (2.8) by "nicer" patterns E^R that satisfy the relationship:

$$\mathbf{E^R} = \mathbf{E} * \mathbf{R} \tag{2.27}$$

where the $K \times K$ matrix $\mathbf{R}$ is chosen such that the resulting "rotated" patterns E^R maximize a certain "simplicity function". Richman (1986) gives

some criteria for "simple" patterns and proposes several forms of simplicity functions. The rotated EOFs constitute a new eigenvector basis and span the same space as the original EOFs. The rotation is called orthogonal if matrix **R** is orthonormal, and oblique otherwise. The most widely used method is the orthogonal varimax rotation (Richman, 1986). The rotated EOFs have none of the properties of the EOFs: they are not the most efficient at explaining the variance of the data, their principal components are no longer orthogonal, and, if the rotation is oblique, the spatial patterns are not orthogonal either. However, the rotated EOFs are often better descriptors of the physical mechanisms underlying the data set.

The rotation of EOFs is a rather controversial topic and the opinion of the community is divided about it. Some defend the use of rotation as a way of finding physically meaningful and more stable patterns. Others are less convinced due to the arbitrariness in the definition of the simplicity function. As a general rule, if the EOF patterns are to be used for purposes other than a physical interpretation (that is, prediction, pattern recognitions, noise reduction, etc), a rotation is generally not necessary. If the patterns are to be physically interpreted, the rotation may be useful in some cases.

Examples and further reading on rotation of EOFs can be found in Lanzante (1984); Richman (1986); Barnston and Livezey (1987); Chelliah and Arkin (1992); Houghton and Tourre (1992); Cheng et al. (1995); Tourre and White (1995); von Storch (1999); Mestas-Nuñez and Enfield (1999) and von Storch and Zwiers (1999).

2.1.8 EOF and Spectral Analysis

A commonly used approach to signal detection in time and space is based on the application of EOF analysis or some of its variations (Combined EOF, SVD, CCA, see sections below) followed by a spectral analysis of the resulting principal components of the modes of interest. The spectral representation of these time series may provide information about the dominant timescales associated with each EOF mode. However, a principal component does not necessarily have a dominant timescale, since the EOF decomposition is explicitly designed to optimally separate spatial patterns and *not* frequencies or timescales. Hence, the principal component associated to a given EOF mode cannot be *expected* to show variability in a preferred frequency band. Nevertheless, it is often the case that climatic signals tend to have structures that are coherent in space and frequency at the same time. That means that a given spatial pattern is often associated to a more or less characteristic timescale of oscillation, in which case the combination of the EOF and spectral analysis may be of use. The Multi Taper Method (MTM) of spectral analysis described in section 3.1 is well suited for this application. Some examples of this com-

bination of techniques can be found in Trenberth and Shin (1984); Deser and Blackmon (1993); Tanimoto et al. (1993) and Venegas et al. (1996).

2.1.9 Examples of EOF analysis on synthetic data sets

Five synthetic data sets are constructed in order to test the behavior of the EOF technique described above. The multivariate (space-time) data sets consist of time series of 1200 samples over a grid of 21×21 points (441 time series in total). According to the notation used in the text: N=1200 and M=441. The 1200 samples are meant to emulate monthly data over 100 years (1200 months). All five fields are given equal amplitude.

The first data set ("Mon15") consists of a monopole spatial pattern centered in the middle of the domain, which oscillates with a period of 15 years (180 months). Its spatial structure is given in Figure 2.2 together with its temporal evolution at point (e).

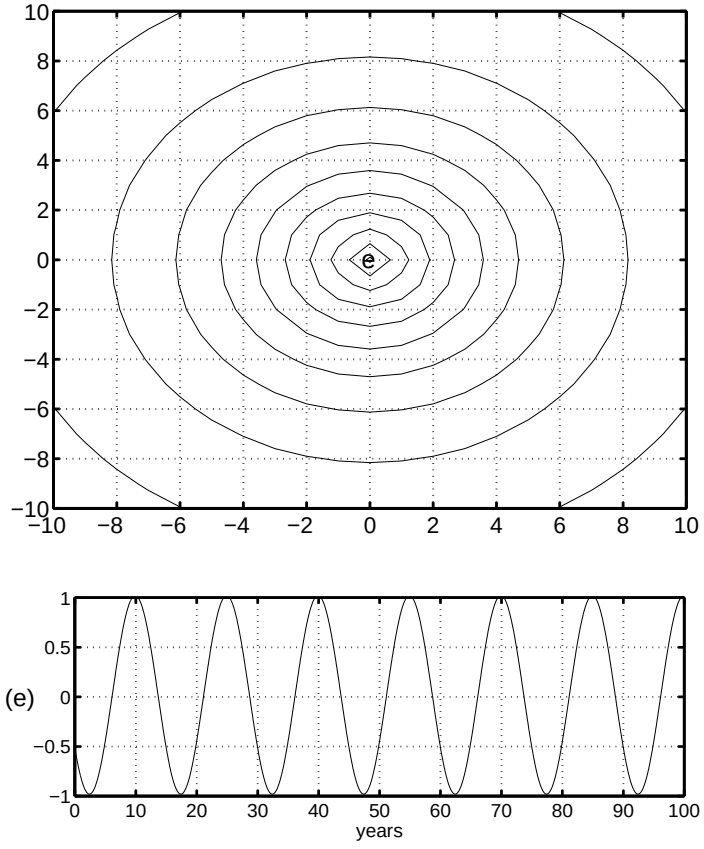

Figure 2.2: Spatial and temporal structure of data set "Mon15"

The second data set ("Dip5") has a north-south dipole pattern in space and oscillates with a period of 5 years (60 months) Figure 2.3 shows its spatial structure and the associated time series at points (a) and (b).

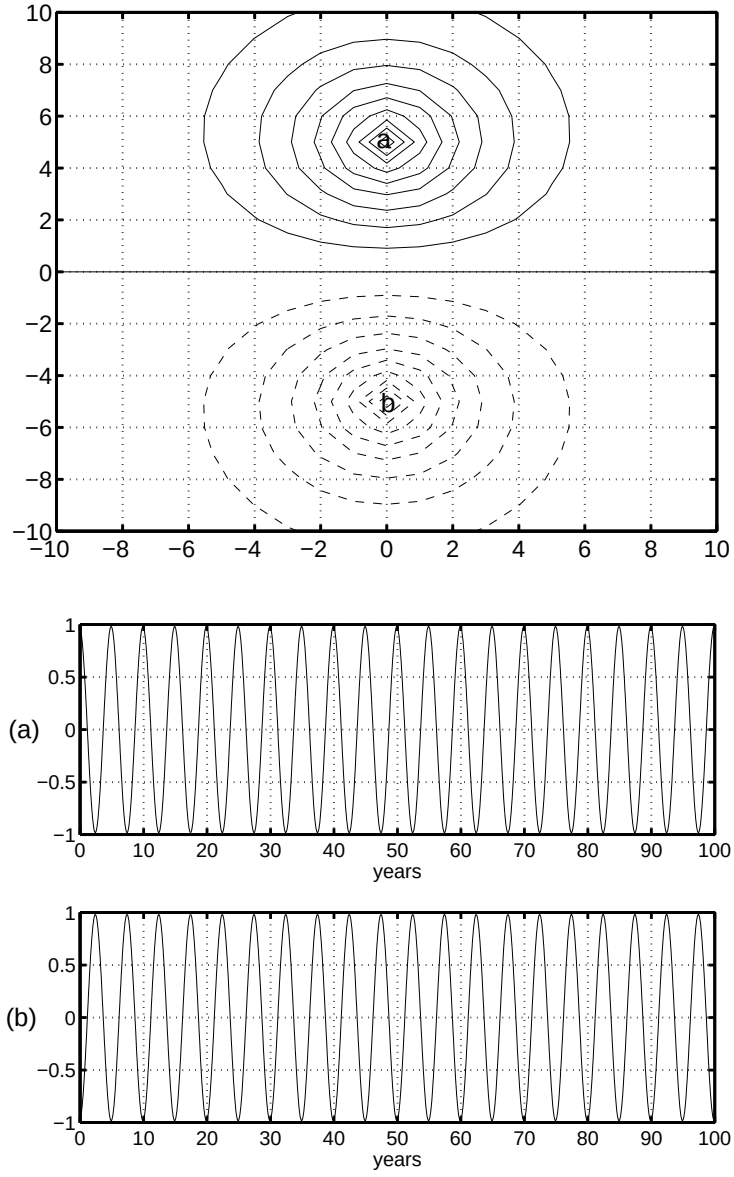

Figure 2.3: Spatial and temporal structure of data set "Dip5"

The third data set ("Dip15") has an east-west dipole pattern in space and oscillates with a period of 15 years (180 months). Figure 2.4 shows its spatial structure and the corresponding time series at points (c) and (d).

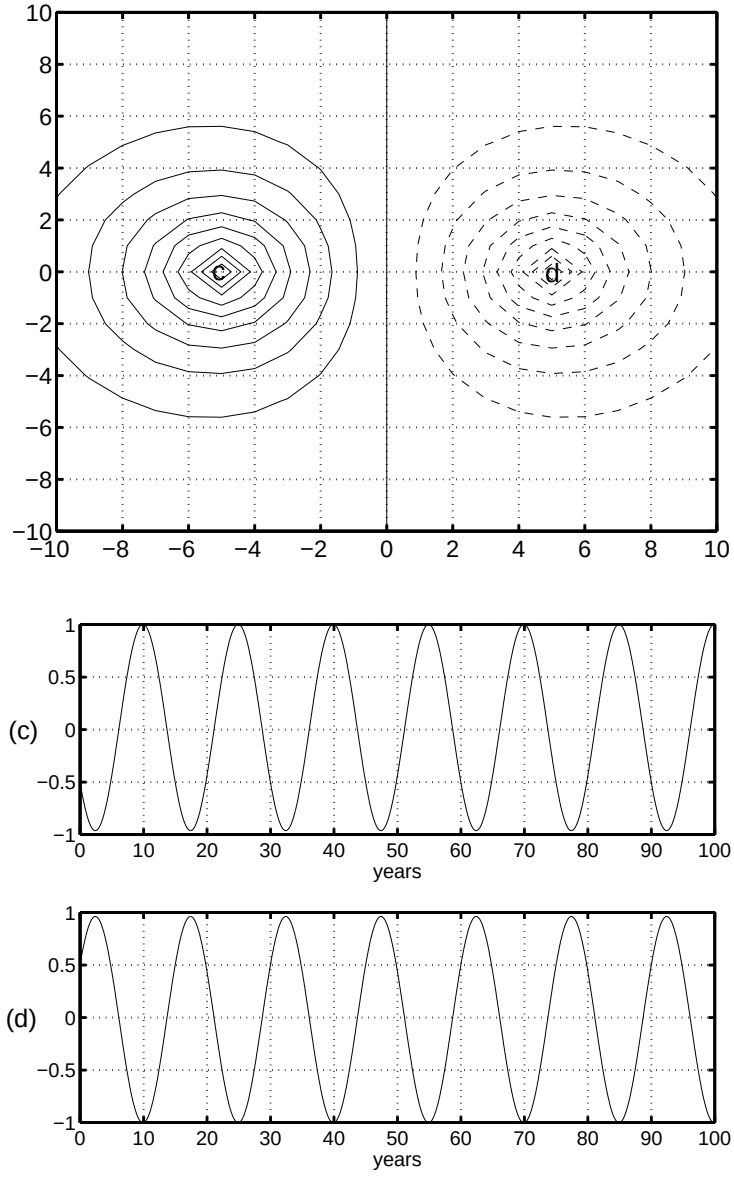

Figure 2.4: Spatial and temporal structure of data set "Dip15"

The fourth data set ("Tri30") has three centers of action in the zonal direction and oscillates with a period of 30 years (360 months). Figure 2.5 shows its spatial structure with the associated time series at points (e) and (f).

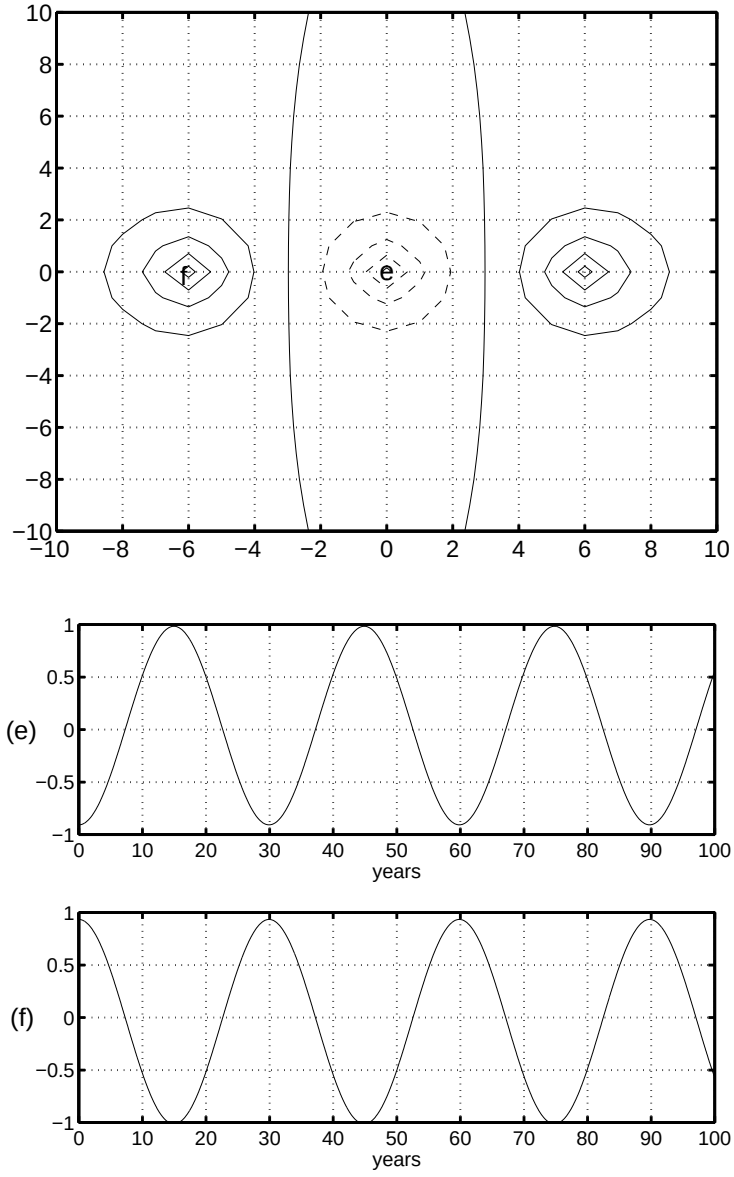

Figure 2.5: Spatial and temporal structure of data set "Tri30"

The first four data sets represent standing oscillations. The fifth data set ("Pro10") represents a propagating pattern: two centers of action of opposite sign that propagate in circles around the domain, with a period of 10 years (120 months). Its spatial structure is shown in Figure 2.6 together with its temporal evolution at points (a), (c), (b) and (d), following the direction of propagation.

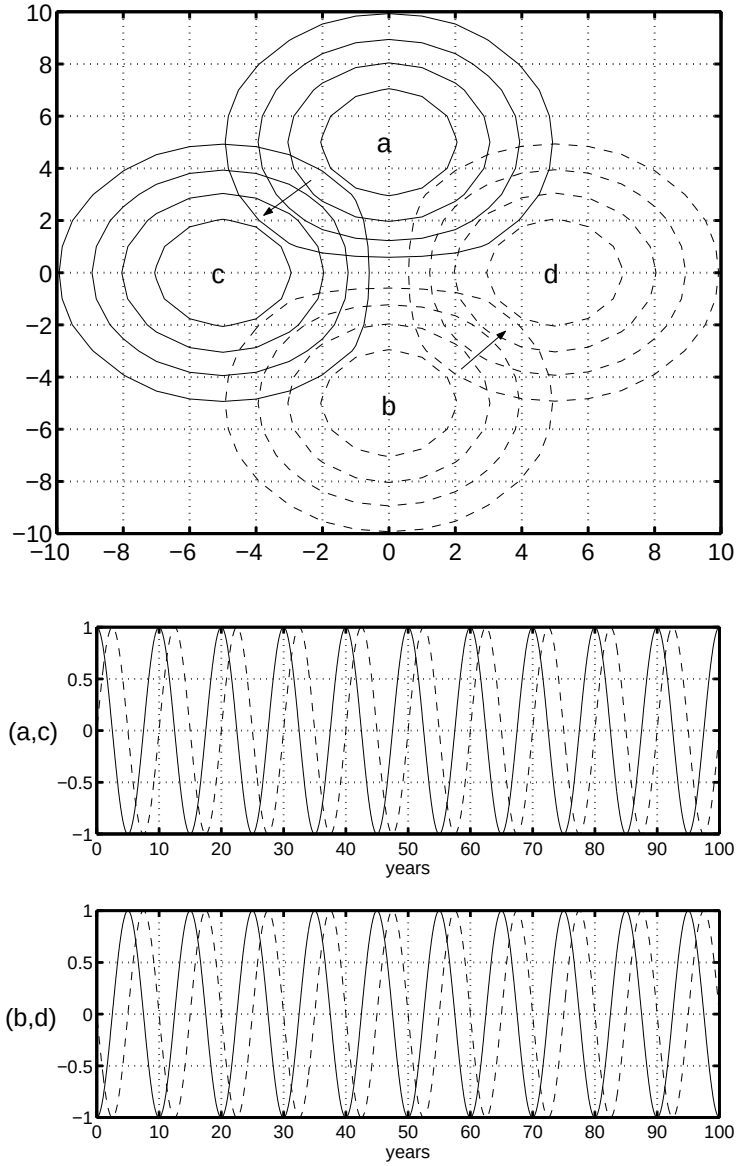

Figure 2.6: Spatial and temporal structure of data set "Pro10"

By linear combinations of two or more of the mentioned synthetic data sets, we will construct several different fields to analyze with the EOF technique. Eight examples of EOF analyses are given below.

Example 1: Field = Dip15 + 2 * Dip5. The two first modes of the EOF decomposition are shown in Figure 2.7. The method isolated Dip5 as the first mode (80% of the total variance), and Dip15 as the second mode (20% of the total variance). Dip5 was clearly dominant since its amplitude was multiplied by 2 in the definition of the field.

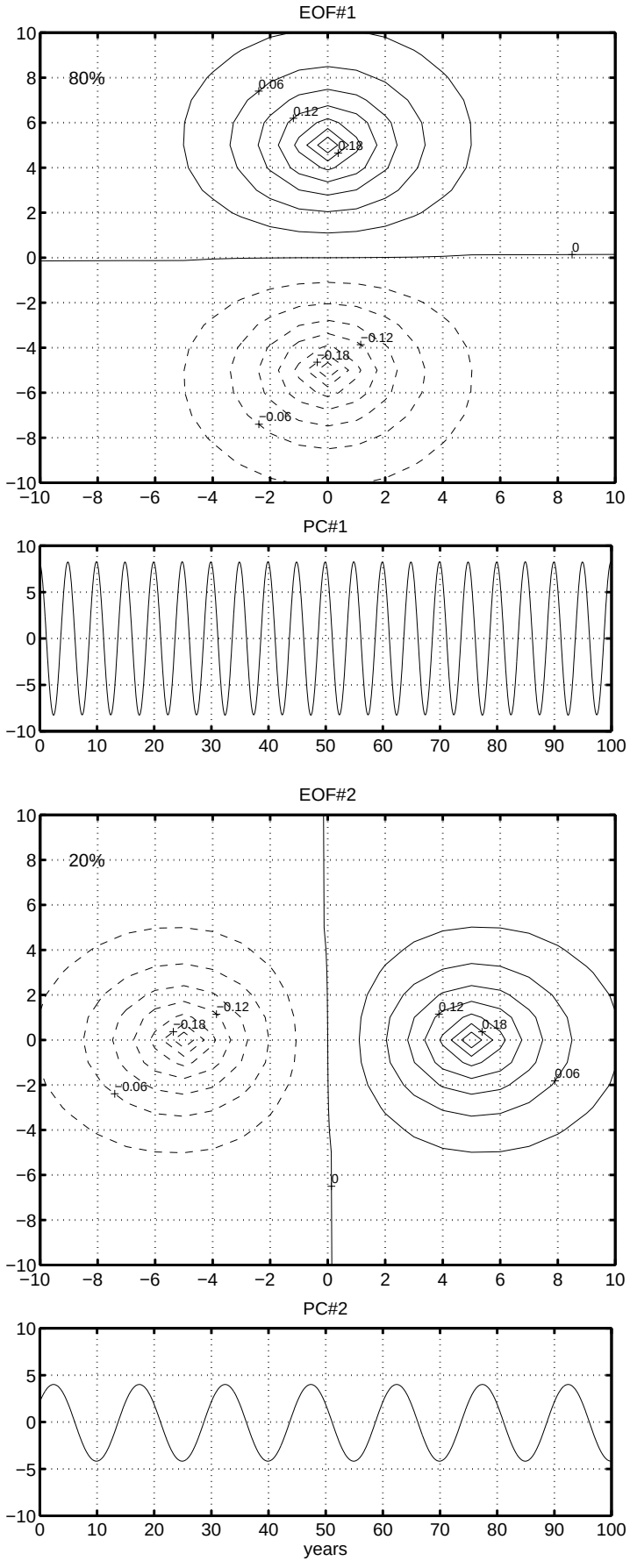

Figure 2.7: EOF decomposition of Example 1. Field = Dip15 + 2 * Dip5.

Example 2: Field = Dip15 + Dip5. The two first modes are shown in Figure 2.8. In this case, since both patterns are dipoles comparable in amplitude, the method is not able to isolate the two expected structures. Instead, it decomposes the data into two modes that equally explain the total variance (50% each), and whose spatial and temporal patterns are a mixture of Dip5 and Dip15. The method is not able to separate the different timescales as we would have liked in this case.

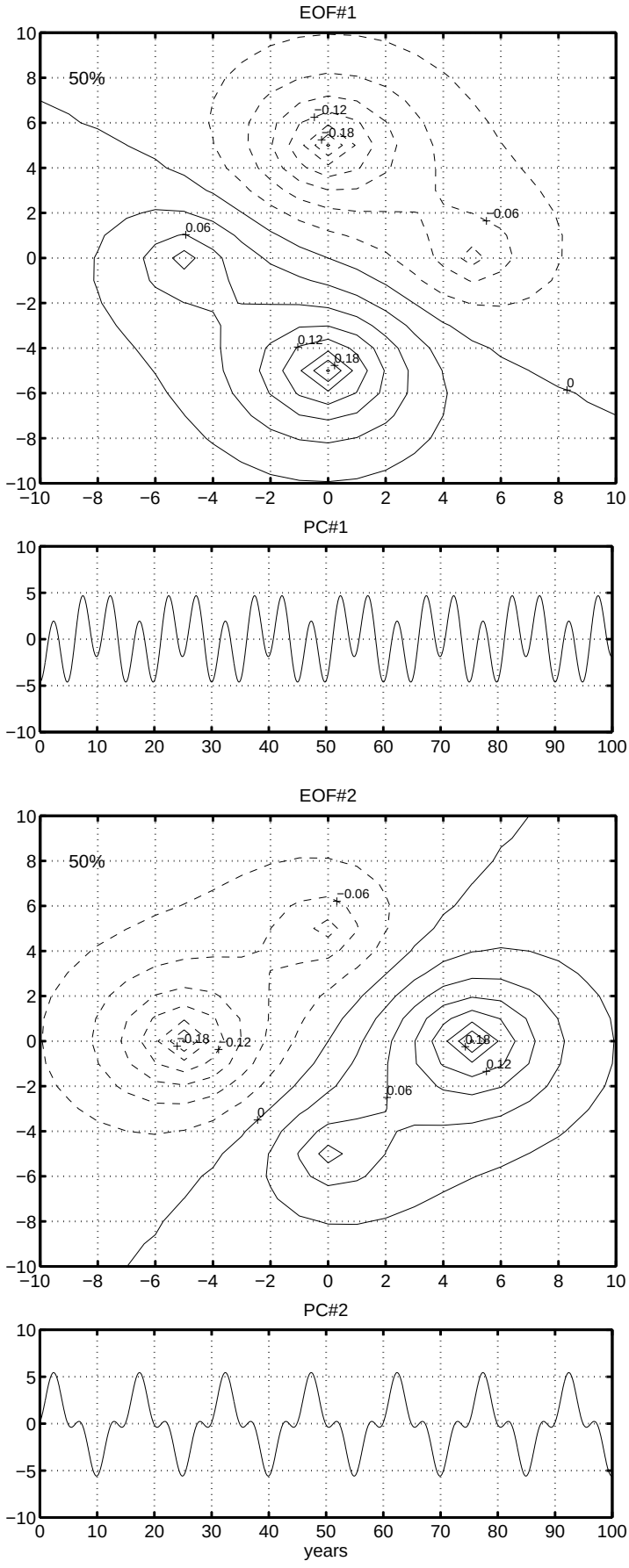

Figure 2.8: EOF decomposition of Example 2. Field = Dip15 + Dip5.

Example 3: Field = Mon15 + Dip5. The two first modes are shown in Figure 2.9. In contrast to what occurred in Example 2, in this case the method clearly isolates Mon15 as the first mode (69%) and Dip5 as the second mode (31%). The difference with Example 2 is that, even though both fields are of comparable amplitude, one of them is a monopole pattern and the other a dipole. This illustrates the fact that the EOF decomposition will always tend to isolate as first mode the structure with largest spatial scale, in this case, the monopole structure Mon15.

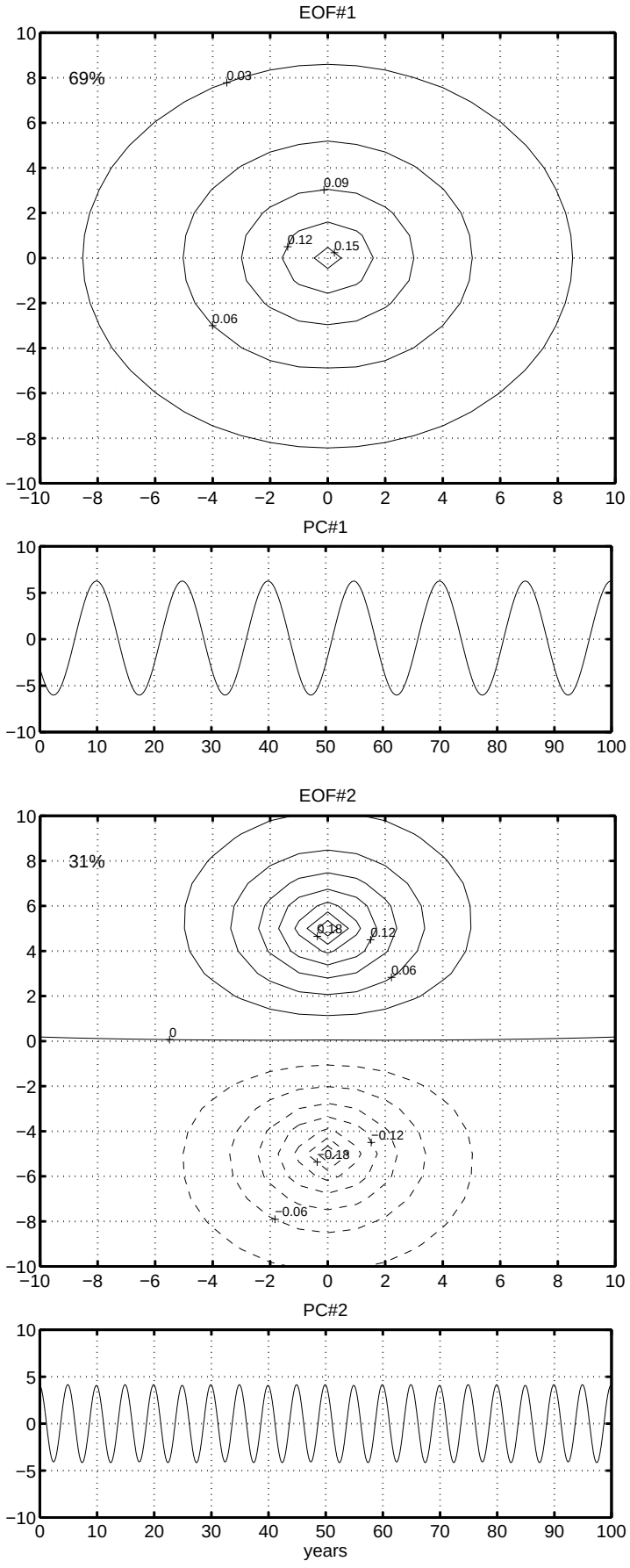

Figure 2.9: EOF decomposition of Example 3. Field = Mon15 + Dip5.

Example 4: Field = Mon15 + 2 * Dip5. The two first modes are shown in Figure 2.10. The fact of multiplying Dip5 by 2 in the definition of the field of Example 3 results in a different decomposition: now Dip5 is the dominant mode and Mon15 is the second.

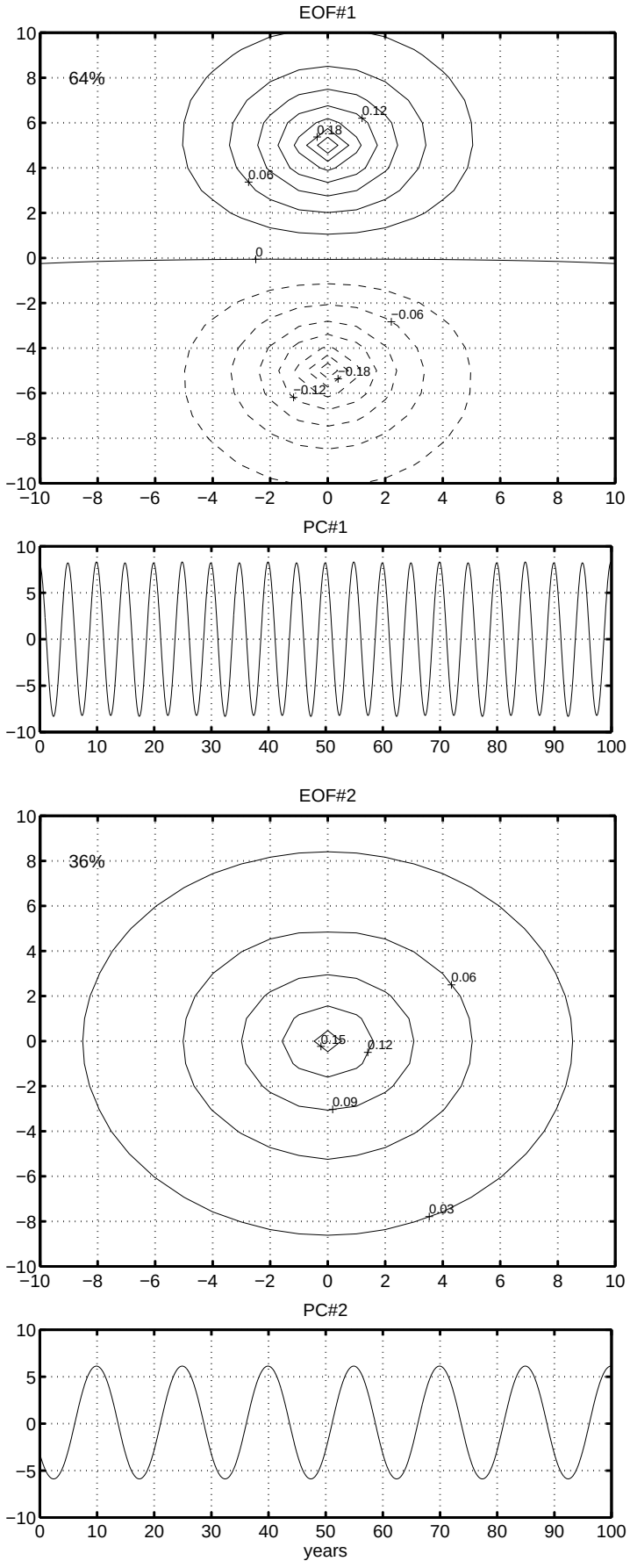

Figure 2.10: EOF decomposition of Example 4. Field = Mon15 + 2 * Dip5.

Example 5: Field = Pro10. The two first modes are shown in Figure 2.11. This example illustrates the fact that the traditional EOF technique is not able to detect the propagation of structures in Pro10, although it may give an indication of its existence. The two first modes in this decomposition are in quadrature (90◦ out of phase) in both space and time. This is an attempt of the method to represent the propagation using standing patterns. One can imagine that the oscillating signal starts with the pattern of the first mode, then a quarter of a cycle (90◦) later it becomes the pattern of the second mode, a quarter of a cycle later it becomes the pattern of the first mode with opposite sign, and a quarter of a cycle later it becomes the pattern of the second mode with opposite sign, which completes a cycle. The percentages of the variance explained are of comparable magnitude (26% and 25%) and the higher order modes (not shown) try to represent the intermediate stages of the propagating signal, always by pairs of quadrature signals and by means of standing patterns.

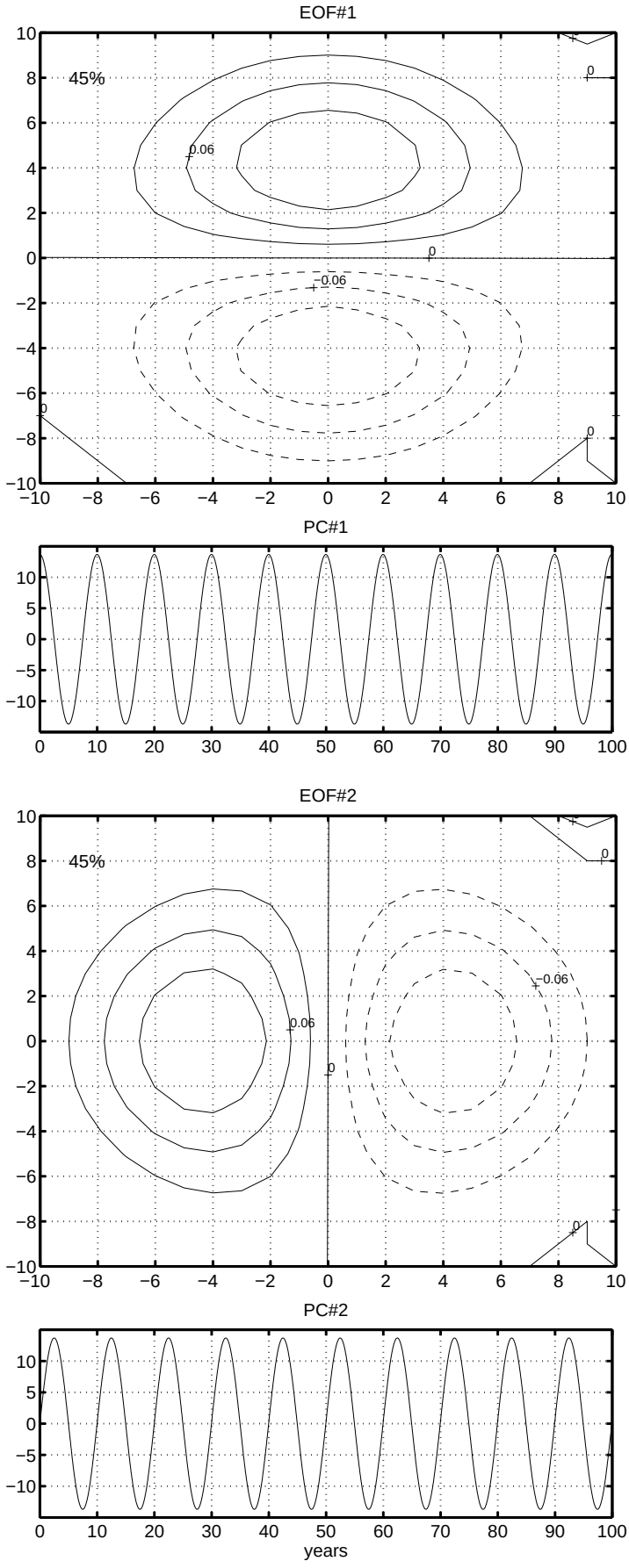

Figure 2.11: EOF decomposition of Example 5. Field = Pro10.

Example 6: Field = Mon15 + Dip5 + Tri30. The three first modes are shown in Figures 2.12 and 2.13. The first mode clearly isolates the monopole pattern Mon15 (58%), the second mode isolates the dipole Dip5 (26%) and the third isolates the three centers of Tri30 (16%). As in example 3, this example illustrates again the fact that the EOF method will choose as first modes those patterns with the largest number of grid points with anomalies of the same sign, that is, it will "prefer" the monopole pattern to the dipole, and the dipole to the three centers, even though all three patterns have comparable amplitude by definition.

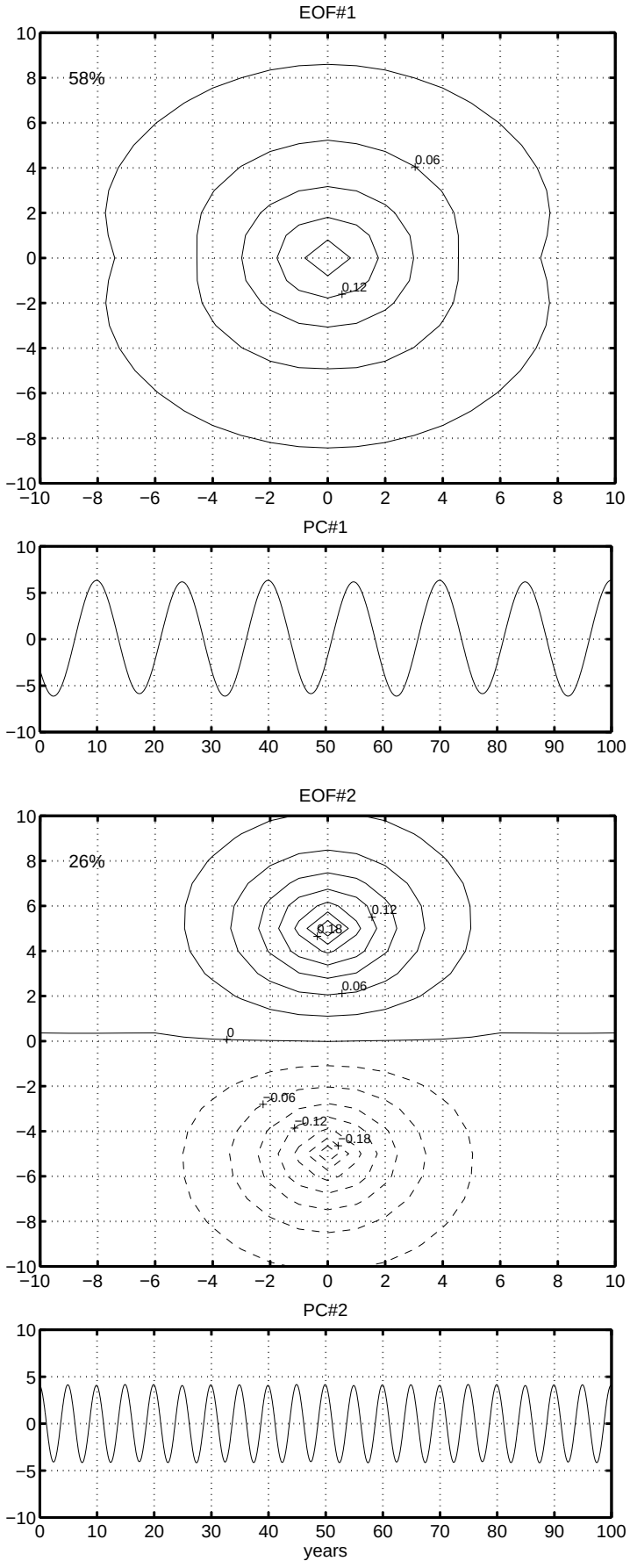

Figure 2.12: EOF decomposition of Example 6. Field = Mon15 + Dip5 + Tri30.

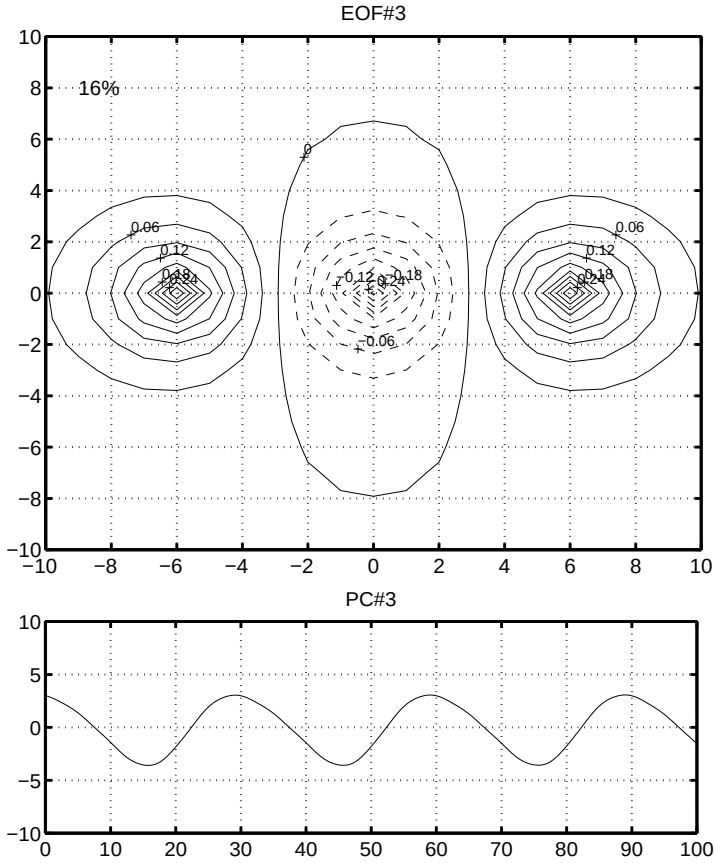

Figure 2.13: EOF decomposition of Example 6. Field = Mon15 + Dip5 + Tri30 (cont'd).

Example 7: Field = Dip5 + Dip15 + Pro10. The two first modes are shown in Figure 2.14. The first and second modes explain a comparable fraction of the variance (46% and 43%) and their spatial patterns seem to isolate the dipole structures of Dip5 and Dip15. However, the temporal evolution of mode 1 shows a mixture of the Dip5 and Pr010 frequencies (5 and 10 years), and that of mode 2 shows a mixture of the Dip15 and Pro10 frequencies (15 and 10 years). That is, the structure of the propagating pattern Pro10 is blended together with the two dipoles. The method is not able to correctly isolate the three patterns in this case, since the Pro10 spatial pattern resembles the two dipoles in certain stages of its propagation.

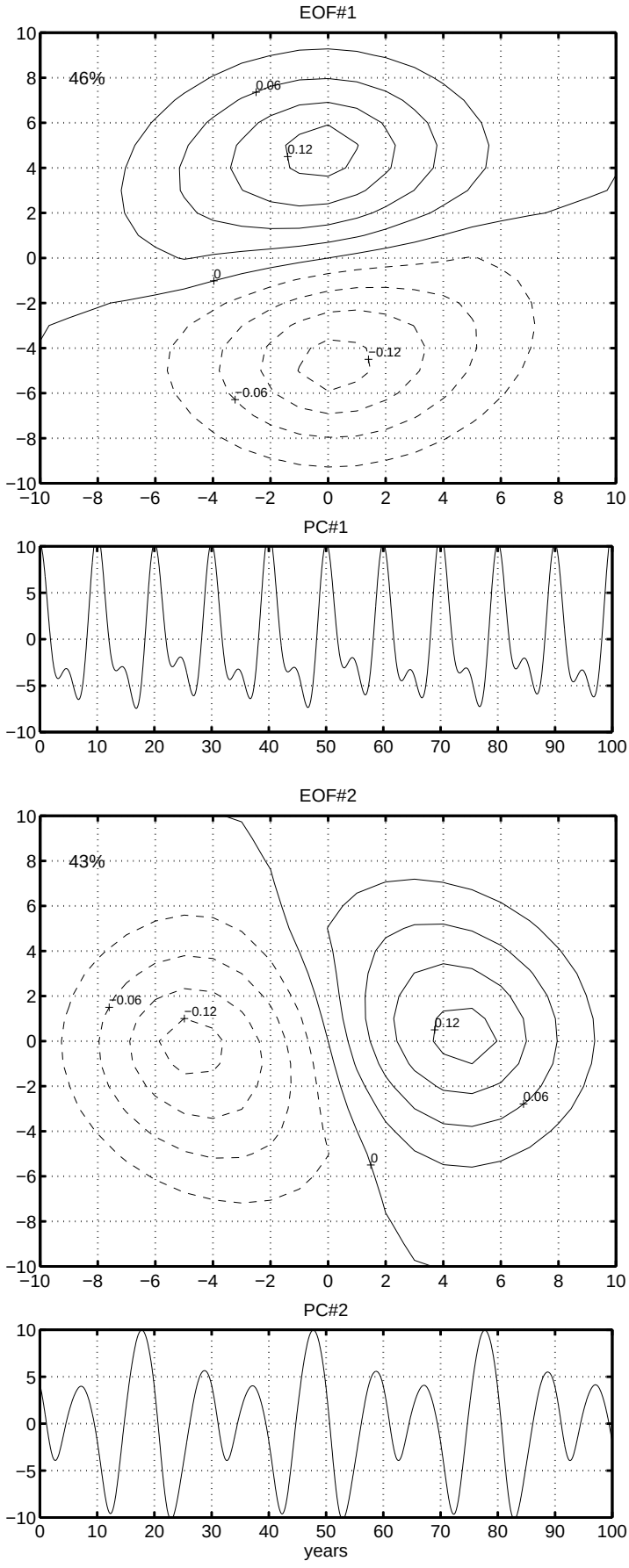

Figure 2.14: EOF decomposition of Example 7. Field = Dip5 + Dip15 + Pro10.

Example 8: Field = Mon15 + Dip5 + Dip15 + Pro10 + Tri30. The four first modes are shown in Figures 2.15 and 2.16. The combination of all five patterns in one results in a rather complicated field. Nevertheless, the decomposition is able to detect the main spatial structures present in the data. The two first modes account for similar percentages of the variance. As in Example 7, mode 1 blends together the east-west dipole pattern Dip15 with the propagating pattern Pro10, which is reflected in the mixture of timescales (10 and 15 years) in PC#1. Mode 2 blends together the north-south dipole pattern Dip5 with another phase of the propagating pattern Pro10, also reflected in the mixture of timescales (5 and 10 years) in PC#2. The monopole pattern Mon15 is partially captured in mode 1, and is responsible for the skewed dipole structure of EOF#1. The remaining variance of Mon15 seems to be accounted for by mode 3, but PC#3 shows variability at periods of 10, 15 and 30 years, so in fact mode 3 is not clearly isolating any of the original patterns. Mode 4, however, shows a rather clean picture of the three-centered pattern Tri30, both in space and time.

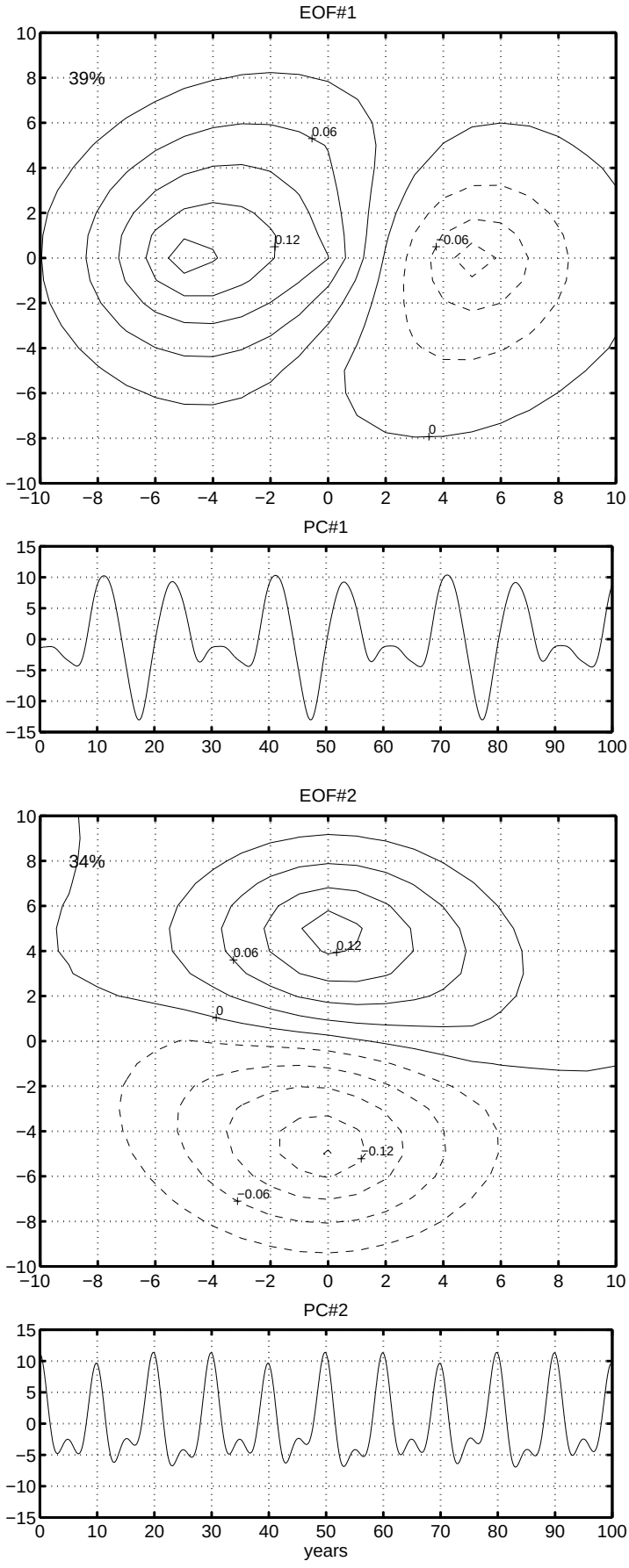

Figure 2.15: EOF decomposition of Example 8. Field = Dip5 + Dip15 + Pro10.

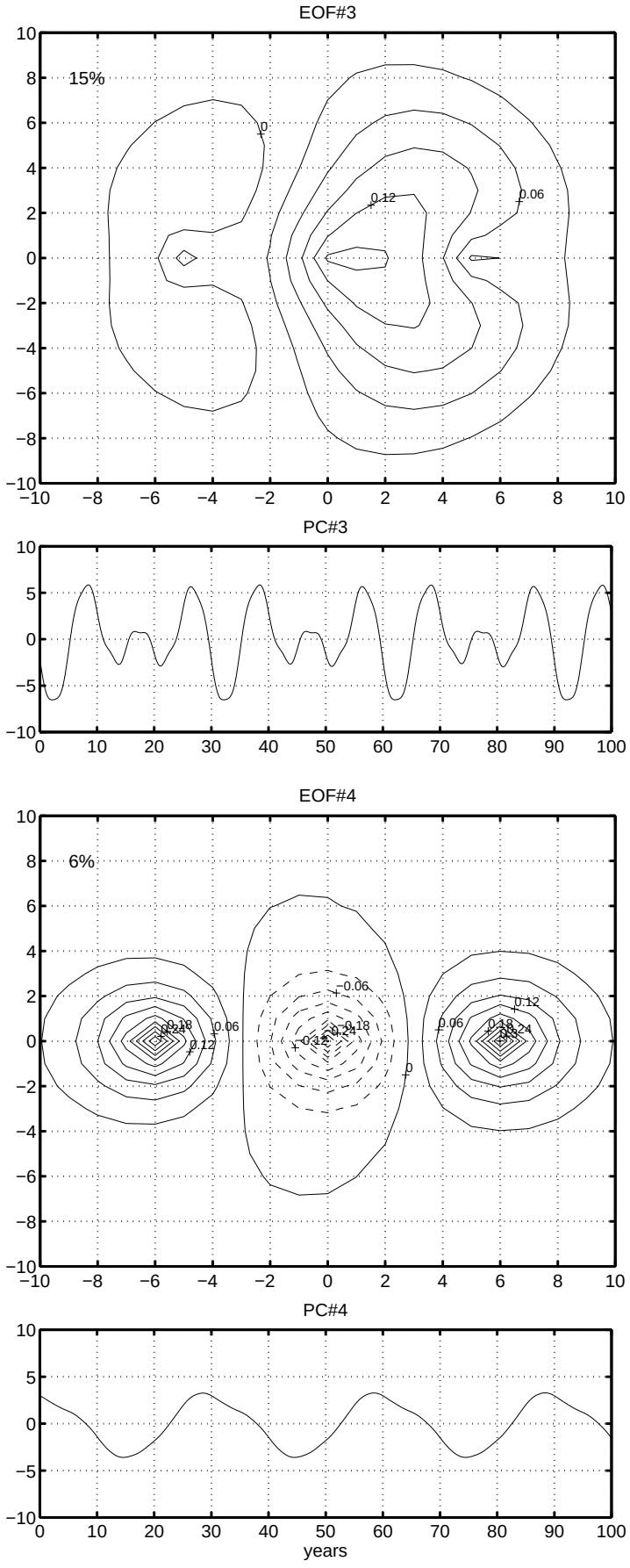

Figure 2.16: EOF decomposition of Example 8. Field = Mon15 + Dip5 + Dip15 + Pro10 + Tri30 (cont'd).

As we have seen through this section, conventional EOF analysis is limited by a number of factors including the dependence of the solution on the domain of analysis, the requirement for orthogonal spatial patterns and uncorrelated principal components, and the inability for separating variability on different frequency bands. In addition, the technique is able to detect standing waves but not propagating features. Over the years, several investigators have developed extensions/variations of the traditional technique in order to overcome some of its limitations. Most of the variations differ in the way they construct the matrix on which the eigenproblem is to be solved. Most of the following sections give descriptions of a variety of methods that belong to the EOF family or are somehow related to it.

2.2 Combined Empirical Orthogonal Functions

The simplest variation to EOF analysis is the Combined EOF, developed to investigate the covariability (or joint variability) of two or more fields at a time. In Combined EOF analysis, the data matrix $\mathbf{F}$ (matrix (2.4)) is constructed with vectors of two or more variables concatenated after one another. An analysis of two variables is outlined here, which can easily be extended to three or more variables.

Let's assume we have two fields S and P we want to combine in one analysis in order to study their joint variability. Mathematically, any combination of scalar fields is permitted, since this is a statistical analysis and not a dynamical analysis. It is the responsibility of the researcher to choose the appropriate fields in order to obtain physically relevant results. The combined fields might be, for example, one atmospheric variable (sea level pressure) and one oceanic variable (sea surface temperature), which may result in a relevant analysis of atmosphere-ocean interaction or coupled variability. Field S consists of time series of length N measured at locations $m = 1 \ldots M_s$. Field P consists of time series of length N measured at locations $m = 1 \ldots M_p$. The number of locations M_s and M_p need not be the same (nor the locations themselves need to be the same for the two fields). However, the length of all the series, N, must be equal for the two variables. The record means are also removed from time series S and P, as was done for field F in section 2.1. As we mentioned in section 2.1.1, the normalization of the fields by their own standard deviation is extremely important in Combined EOF analysis, in order to avoid the dominance of one field over the other. For example, if air pressure from midlatitudes is analyzed together with air pressure from low latitudes without normalizing the fields, then the resulting EOFs will be mostly influenced by the high-variance mid-latitude areas. If a vector is made up by combining

unnormalized data of temperature in units of K and precipitation in units of m/s^8, the EOF patterns will concentrate on the temperature entries, simply due to their larger magnitude. An extra weighting of the data matrix may be made if the number of locations in one field largely exceeds the number of locations in the other ($M_s \gg M_p$ or vice-versa). In this case, the values of each field should be divided by the total number of time series (locations) available for that field: S/M_s and P/M_p, thereby avoiding the dominance of one field over the other due to sampling differences.

The aggregated data matrix F is then constructed as follows:

$$\mathbf{F} = \begin{bmatrix} S_1(1) & S_1(2) & \dots & S_1(N) \\ S_2(1) & S_2(2) & \dots & S_2(N) \\ \dots & \dots & \dots & \dots \\ S_{M_s}(1) & S_{M_s}(2) & \dots & S_{M_s}(N) \\ P_1(1) & P_1(2) & \dots & P_1(N) \\ P_2(1) & P_2(2) & \dots & P_2(N) \\ \dots & \dots & \dots & \dots \\ P_{M_p}(1) & P_{M_p}(2) & \dots & P_{M_p}(N) \end{bmatrix} \tag{2.28}$$

where rows $1 \dots M_s$ contain the time series of field S, and rows $M_s{+}1 \dots M_s{+}M_p$ contain the time series of field P. All rows have length N. Data matrix $\mathbf{F}$ has dimension $M_s{+}M_p \times N$.

The covariance matrix $\mathbf{R_{FF}} = \mathbf{F} * \mathbf{F}^\dagger$, of dimension $M_s{+}M_p \times M_s{+}M_p$, then results:

$$\mathbf{R_{FF}} = \begin{bmatrix} \langle S_1 S_1 \rangle & \dots & \langle S_1 S_{M_s} \rangle & \langle S_1 P_1 \rangle & \dots & \langle S_1 P_{M_p} \rangle \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \langle S_{M_s} S_1 \rangle & \dots & \langle S_{M_s} S_{M_s} \rangle & \langle S_{M_s} P_1 \rangle & \dots & \langle S_{M_s} P_{M_p} \rangle \\ \langle P_1 S_1 \rangle & \dots & \langle P_1 S_{M_s} \rangle & \langle P_1 P_1 \rangle & \dots & \langle P_1 P_{M_p} \rangle \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \langle P_{M_p} S_1 \rangle & \dots & \langle P_{M_p} S_{M_s} \rangle & \langle P_{M_p} P_1 \rangle & \dots & \langle P_{M_p} P_{M_p} \rangle \end{bmatrix} \tag{2.29}$$

Matrix $\mathbf{R_{FF}}$ contains the spatial covariances between all possible combinations of fields at pairs of locations i, j: $\langle S_i S_j \rangle$, $\langle S_i P_j \rangle$, $\langle P_i S_j \rangle$ and $\langle P_i P_j \rangle$.

As in conventional EOF analysis, we now solve the eigenproblem on this matrix $\mathbf{R_{FF}}$ as in equation (2.8):

$$\mathbf{E} * \mathbf{R_{FF}} = \mathbf{\Lambda} * \mathbf{E} \tag{2.30}$$

The eigenvalues in matrix $\mathbf{\Lambda}$ now measure the fraction of the covariance (or joint variance) between the two fields S and P accounted for by each Combined EOF mode. The resulting eigenvectors E^k in the columns of matrix $\mathbf{E}$ have now length M_s+M_p:

$$\mathbf{E} = \begin{bmatrix} E_1^1 & E_1^2 & \dots & E_1^K \\ E_2^1 & E_2^2 & \dots & E_2^K \\ \dots & \dots & \dots & \dots \\ E_{M_s}^1 & E_{M_s}^2 & \dots & E_{M_s}^K \\ E_{M_s+1}^1 & E_{M_s+1}^2 & \dots & E_{M_s+1}^K \\ E_{M_s+2}^1 & E_{M_s+2}^2 & \dots & E_{M_s+2}^K \\ \dots & \dots & \dots & \dots \\ E_{M_s+M_p}^1 & E_{M_s+M_p}^2 & \dots & E_{M_s+M_p}^K \end{bmatrix} \tag{2.31}$$

Each eigenvector includes a spatial pattern of field S in elements $1 \dots M_s$ *and* a spatial pattern of field P in elements $M_s+1 \dots M_s+M_p$. We thus need to divide each eigenvector E^k into two separate EOFs, E_S^k and E_P^k, one for each field. Note that, in matrix (2.31), we have considered only K columns of $\mathbf{E}$, instead of M_s+M_p columns. As in the preceding sections, this is because only the first K eigenvalues in matrix $\mathbf{\Lambda}$ are non zero, with $K \leq min(M_s+M_p, N)$. Hence, we consider only the K eigenvectors that are associated with the non-zero eigenvalues.

The principal components are then computed by projecting each original field, S and P, onto its own eigenvectors, E_S^k and E_P^k, using equations (2.11)-(2.12). Thus, two sets of EOF patterns E_S^k and E_P^k, and two sets of principal components A_S^k and A_P^k are obtained, one set for each of the two variables analyzed jointly.

While this procedure allows for an analysis of two or more fields together, it is clear that it requires even larger computational power than conventional EOF since the size of the covariance matrix increases considerably with the number of fields analyzed. The alternative Singular Value Decomposition (SVD) analysis of coupled fields described in the next section gives a more efficient and robust solution to the problem of joint fields analysis. Applications of Combined EOF analysis can be found for example in Bretherton et al. (1992); Wallace et al. (1992); Deser and Blackmon (1993); and Kousky and Kayano (1994).

2.3 Singular Value Decomposition (SVD)

The Singular Value Decomposition (SVD) of coupled fields allows for the identification of pairs of EOFs and PCs which account for a fraction of the covariance between two variables analyzed jointly. The SVD analysis of coupled fields gives exactly the same results as the Combined EOF analysis of two variables described in section 2.2. The main difference between the two approaches resides on the larger efficiency and robustness of the SVD method.

The name SVD given to this technique is sometimes misleading since it blends together the algebraic solution of a matrix problem (the singular value decomposition described in section 2.1.4) with the problem itself, that is, the maximization of the covariance between two fields. To avoid misunderstandings, in this review we will use the abbreviation "SVD" or the capitalized words "Singular Value Decomposition" when referring to the method of analysis of two fields and just "singular value decomposition" when referring to the algebraic problem. Due to this conflict, some authors have preferred to name this technique Maximum Covariance Analysis (MCA, Bretherton et al., 1992)

After preparing the two datasets of variables S and P as described in section 2.1.1, we organize the time series of each of the two variables *separately* in two data matrices as in matrix (2.4). We thus obtain an $M_s \times N$ data matrix $\mathbf{S}$ and an $M_p \times N$ data matrix $\mathbf{P}$. The spatial dimensions of the two variables, M_s and M_p, need not be the same, but all the time series must have the same number of time steps N.

We then construct the cross-covariance matrix between time series of S and P as $\mathbf{R_{SP}} = \mathbf{S} * \mathbf{P}^\dagger$:

$$\mathbf{R_{SP}} = \begin{bmatrix} \langle S_1 P_1 \rangle & \langle S_1 P_2 \rangle & \dots & \langle S_1 P_{M_p} \rangle \\ \langle S_2 P_1 \rangle & \langle S_2 P_2 \rangle & \dots & \langle S_2 P_{M_p} \rangle \\ \dots & \dots & \dots & \dots \\ \langle S_{M_s} P_1 \rangle & \langle S_{M_s} P_2 \rangle & \dots & \langle S_{M_s} P_{M_p} \rangle \end{bmatrix} \tag{2.32}$$

where each element $\langle S_i P_j \rangle$ is the spatial cross-covariance between time series S_i and P_j (at locations i and j), as defined in equation (2.7). Matrix $\mathbf{R_{SP}}$ is $M_s \times M_p$ and needs not be square, as in general $M_s \neq M_p$. Since we are again comparing two fields, the considerations about normalization and weighting of the two variables mentioned in section 2.2 also apply in this case. Strictly speaking, if the time series of $\mathbf{S}$ and $\mathbf{P}$ are standarized, the cross-covariance matrix $\mathbf{R_{SP}}$ is in fact a cross-correlation matrix.

We now perform a singular value decomposition on $\mathbf{R_{SP}}$, that is, we find matrices $\mathbf{U}$, $\mathbf{\Gamma}$ and $\mathbf{V}$ such that:

$$\mathbf{R_{SP}} = \mathbf{U} * \mathbf{\Gamma} * \mathbf{V}^{\dagger} \tag{2.33}$$

The *columns* of the quadratic $M_s \times M_s$ matrix $\mathbf{U}$ are orthogonal and contain the singular vectors of S. The *rows* of the quadratic $M_p \times M_p$ matrix $\mathbf{V}^{\dagger}$ are also orthogonal and contain the singular vectors of P. Diagonal matrix $\mathbf{\Gamma}$ is rectangular $M_s \times M_p$ with zero elements outside the diagonal and positive or zero elements on the diagonal. The scalars on the diagonal, γ_k, are the singular values and are placed in decreasing order of magnitude. There are only $K \leq min(M_s, M_p)$ non-zero singular values in $\mathbf{\Gamma}$, which defines the maximum number of SVD modes we can determine. This implies that there are only K useful singular vectors for each variable and hence the effective dimension of matrix $\mathbf{U}$ is $M_s \times K$ and of matrix $\mathbf{V}^{\dagger}$ is $K \times M_p$.

The principal components of fields S and P are obtained by projecting each field onto its respective singular vectors as in equations (2.11)- (2.12). For field S:

$$\mathbf{A} = \mathbf{U}^{\dagger} * \mathbf{S} \tag{2.34}$$

$$A^k(t) = \sum_{m=1}^{M_s} U_m^k \, S_m(t) \tag{2.35}$$

And for field P:

$$\mathbf{B} = \mathbf{V}^{\dagger} * \mathbf{P} \tag{2.36}$$

$$B^k(t) = \sum_{m=1}^{M_p} V_m^k \, P_m(t) \tag{2.37}$$

$\mathbf{A}$ is $K \times N$ and its rows contain the principal components of field S. $\mathbf{B}$ is $K \times N$ and its rows contain the principal components of field P.

The K non-zero singular values γ_k are proportional to the squared covariance fraction between fields S and P accounted for by the mode k, such that:

$$\%\ \text{Squared Covariance Mode k} = \frac{\gamma_k^2}{\sum_{i=1}^{K} \gamma_i^2} * 100 \tag{2.38}$$

Each SVD mode of covariability between S and P is thus determined by a pair of spatial patterns (one for each field), a pair of principal components that show how the respective spatial patterns evolve in time, and a singular value that indicates how much of the squared covariance between the two fields is accounted for by the mode. The correlation coefficient between the two PCs of mode k, $\mathbf{r}[A^k(t), B^k(t)]$, provides a measure of how strongly fields S and P are related to one another through mode k. It is this quantity that is maximized in Canonical Correlation Analysis discussed in the following section. Further reading on SVD analysis of coupled fields and some applications can be found in Wallace et al. (1992); Bretherton et al. (1992); Cherry (1996); Venegas et al. (1996); Peng and Fyfe (1996); Cherry (1997); Chang et al. (1997); Venegas et al. (1997); Deser and Timlin (1997); and Yi et al. (1999).

Just as in conventional EOF analysis, it is common to present the singular vectors or spatial patterns as correlation maps. In SVD analysis, both "homogeneous" and "heterogeneous" maps can be defined. The kth homogeneous correlation map is constructed as the map of correlation coefficients between the principal component of the kth mode of a field and the values of the *same* field at each grid point, that is, $\mathbf{r}[A^k(t), S(t)]$ or $\mathbf{r}[B^k(t), P(t)]$. It is useful as an indicator of the spatial localization of the covarying part of the field and its kth mode. The kth heterogeneous correlation map is constructed as the map of correlation coefficients between the principal component of the kth mode of a field and the values of the *other* field at each grid point, that is, $\mathbf{r}[A^k(t), P(t)]$ or $\mathbf{r}[B^k(t), S(t)]$. It indicates how well the grid point values of one field can be predicted from the knowledge of the principal component of the other. The singular vector, the homogeneous and the heterogeneous map of a given field generally present almost identical distributions of centers of action. It is expected that the heterogeneous correlations tend to be weaker than the homogeneous ones. Differences in the shapes of the patterns on the homogeneous and heterogeneous maps for the same field, if significant, can sometimes provide information about the nature of the cause/effect links between the fields. Examples of correlation maps can be found in Barnett and Preisendorfer (1987); Wallace et al. (1992); Frankignoul et al. (1996); and Venegas et al. (1997).

2.4 Canonical Correlation Analysis (CCA)

The Canonical Correlation Analysis (CCA) technique attempts to find a linear relationship between fields S and P by maximizing the correlation coefficient between them, instead of maximizing the squared covariance as in the SVD technique discussed in section 2.3. The CCA isolates the linear combination of data of field S and the linear combination of data of field P that have maximum

correlation coefficient. This pair of time series is more strongly correlated than the principal components of the leading pair of patterns obtained from the SVD analysis of section 2.3, but explains a smaller fraction of the covariance between the two fields.

We start as usual by preparing the two simultaneously observed fields S and P in the way shown in section 2.1.1 and by constructing data matrices $\mathbf{S}$ and $\mathbf{P}$ of dimension $M_s \times N$ and $M_p \times N$, respectively, as in equation (2.4). The next step is to form three covariance matrices: $\mathbf{R_{SS}} = \mathbf{S} * \mathbf{S}^\dagger$ is the covariance matrix of field S computed as in equation (2.6), $\mathbf{R_{PP}} = \mathbf{P} * \mathbf{P}^\dagger$ is the covariance matrix of field P computed as in equation (2.6), and $\mathbf{R_{SP}} = \mathbf{S} * \mathbf{P}^\dagger$ is the cross-covariance matrix between fields S and P computed as in equation (2.32). We then form matrices $\mathbf{Q_S}$ and $\mathbf{Q_P}$ as combinations of the three such as:

$$\mathbf{Q_S} = \mathbf{R_{SS}^{-1}} * \mathbf{R_{SP}} * \mathbf{R_{PP}^{-1}} * \mathbf{R_{SP}^\dagger} \tag{2.39}$$

$$\mathbf{Q_P} = \mathbf{R_{PP}^{-1}} * \mathbf{R_{SP}^\dagger} * \mathbf{R_{SS}^{-1}} * \mathbf{R_{SP}} \tag{2.40}$$

where $\mathbf{R^{-1}}$ is defined as the inverse of the square matrix $\mathbf{R}$. Matrix $\mathbf{Q_S}$ is of dimension $M_s \times M_s$, and matrix $\mathbf{Q_P}$ is of dimension $M_p \times M_p$.

Then we solve the eigenproblem on each of these two rather complicated-looking matrices using equation (2.8).

$$\mathbf{Q_S} * \mathbf{\Pi_S} = \mathbf{\Pi_S} * \mathbf{\Lambda} \tag{2.41}$$

$$\mathbf{Q_P} * \mathbf{\Pi_P} = \mathbf{\Pi_P} * \mathbf{\Lambda} \tag{2.42}$$

The two matrices $\mathbf{Q_S}$ and $\mathbf{Q_P}$ share the same non-zero eigenvalues in matrix $\mathbf{\Lambda}$. For each matrix we obtain a set of eigenvectors called "adjoint patterns": Π_S^k and Π_P^k, where $k = 1 \ldots K$ indicates the mode, as usual. These adjoint patterns are found in the columns of matrices Π_S and Π_P, of dimension $M_s \times M_s$ and $M_p \times M_p$, respectively. Then, the spatial "Canonical Correlation Patterns", (CCP), E_S^k and E_P^k are the columns of matrices E_S and E_P, derived from the adjoint patterns as:

$$\mathbf{E_S} = \mathbf{R_{SS}} * \mathbf{\Pi_S} \tag{2.43}$$

$$\mathbf{E_P} = \mathbf{R_{PP}} * \mathbf{\Pi_P} \tag{2.44}$$

The temporal "Canonical Correlation Coefficients", (CCC), A_S^k and A_P^k, $k = 1 \ldots K$, are the columns of matrices A_S and A_P, also derived from the adjoint patterns such as:

$$\mathbf{A_S} = \mathbf{S}^\dagger * \mathbf{\Pi_S} \tag{2.45}$$

$$\mathbf{A_P} = \mathbf{P}^\dagger * \mathbf{\Pi_P} \tag{2.46}$$

The CCC are constrained to be temporally uncorrelated. However, the CCP are not necessarily orthogonal in space. According to von Storch (1999), the CCC satisfy the following requirements:

1. The correlations between A_S^l and A_S^k, between A_P^l and A_P^k and between A_S^l and A_P^k are zero for all $l \neq k$, where l, k indicate the mode.

2. The correlation between A_S^1 and A_P^1 is maximum and equal to the largest eigenvalue of matrices $\mathbf{Q_S}$ and $\mathbf{Q_P}$.

3. The correlation between A_S^2 and A_P^2 is maximum under the constraints in 1) and 2) and equal to the second largest eigenvalue of matrices $\mathbf{Q_S}$ and $\mathbf{Q_P}$.

4. The correlations between the higher indexed pairs of coefficients satisfy similar constraints, that is, they are maximum while being independent of all previously determined coefficients.

The spatial dimensions M_s and M_p of fields S and P are in general different. As we said, matrix $\mathbf{Q_S}$ is $M_s \times M_s$ and matrix $\mathbf{Q_P}$ is $M_p \times M_p$. If one of the two dimensions M_s or M_p is much bigger than the other, it is recommended to solve the eigenproblem only on the smallest of the two matrices $\mathbf{Q_S}$ or $\mathbf{Q_P}$. As such, if Π_S^k is an eigenvector of $\mathbf{Q_S}$ with a non-zero eigenvalue, then $\mathbf{R_{PP}^{-1}} * \mathbf{R_{SP}^\dagger} * \Pi_S^k$ is an eigenvector of $\mathbf{Q_P}$ with the same eigenvalue.

Some authors strongly recommend to compress the data prior to a CCA (Barnett and Preisendorfer, 1987; Bretherton et al., 1992). This is done through a conventional EOF analysis. The two fields are pre-filtered by retaining only the projection of each field onto a subset of its leading EOF patterns before applying CCA. The motivation for this compression is to minimize the danger of misinterpreting the random details of the noise variability within the sample as true correlations. In particular, when the spatial dimension of the fields M_s and M_p is large compared to the number of samples N, it is probable that spuriously high correlations appear due to the many contributions of the

badly sampled noise. The CCA method may emphasize this sample correlations in detriment of the true signal correlations. The noise reduction achieved by pre-filtering the data through conventional EOF renders the resulting CCA modes more stable with respect to sample variability. Another advantage of the data compression through EOF is the reduction in the spatial degrees of freedom of the fields (the spatial dimensions M_s and M_p) which results in the reduction of the size of matrices $\mathbf{Q_S}$ and $\mathbf{Q_P}$. However, it is important to keep in mind that the results may depend on the a priori EOF truncation of the data, so the sensitivity to this should be investigated. Note that the SVD analysis of coupled field described in section 2.3 does not require any a priori compression of the data.

A last note of caution: like in any kind of correlation-based analysis, the identification of a high correlation does not necessarily prove that the two time series are connected through a cause-effect relationship. It may as well be that a third (unknown) variable is controlling the behaviour of the two time series analyzed. As with the preceding techniques, one must thus be cautious when invoking a CCA result to explain relationships between variables. Further reading and applications of CCA can be found in Barnett and Preisendorfer (1987); Wallace et al. (1992); Bretherton et al. (1992); Zorita et al. (1992); Werner and von Storch (1993); Cherry (1996); von Storch (1999); and von Storch and Zwiers (1999).

Chapter 3

Patterns in Time: Time Series Analysis

3.1 The Multi-Taper Method (MTM)

In contrast to the techniques described in the last chapter, the Multi-Taper Method (MTM) of spectral analysis is a univariate approach, that is, a time series analysis method used for the description of a single time series. Time series analysis methods aim to examine a temporal sequence of data (a time series) in terms of its frequency content. They provide insight on the different types of signals that are blended together and "hidden" inside a noisy time series. In other words, the common goal of most time series analysis is to separate the deterministic periodic oscillations in the data from random and aperiodic fluctuations associated with the background noise (that is, unwanted geophysical variability) or with instrumental errors.

A variety of spectral analysis techniques have been widely employed in the analysis of geophysical processes (see, for example, Brillinger, 1981, and the references therein). However, more sophisticated methods have recently been developed which are more faithful in their underlying assumptions to the irregular oscillatory behaviour expected of climatic signals. Among such methods is the MTM approach of spectral analysis, which makes use of multiple orthogonal data tapers to describe structures in time series that are modulated in frequency and amplitude. More importantly, this method provides a spectral estimate with an optimal trade-off between spectral resolution and variance.

Traditionally, it has been a standard procedure to multiply a time series by a data taper (also called data window) before performing a discrete Fourier transform (DFT), in order to reduce spectral leakage (that is, to minimize the bias in the spectral estimate due to leakage of power from a given frequency to its neighbour frequencies (refer to Percival and Walden, 1993, for details on tapering). The product $a(t)F(t)$ is formed for each time step t, where

$F(t),\ t = 1 \ldots N$, is a zero-mean time series and $a(t),\ t = 1 \ldots N$, is a sequence of real-valued constants called a data taper or data window. The goal of the taper is to "force" the time series to go to 0 at both ends. The spectral window associated with a taper (that is, the representation of the taper in the frequency domain) has much smaller side-lobes than the Fejér's kernel. This results in a better (less biased) estimate of the spectrum (see Percival and Walden, 1993, for further details). However, spectral estimates obtained from a tapered series have relatively large variance. This is so because the data at both ends of the time series is discarded by the taper (since the taper goes to 0 as it approaches the ends) and thereby a considerable amount of statistical information is lost during the tapering procedure. Hence, as long as only one single data taper is used, there will be a trade-off between the resistance to spectral leakage and the variance of the spectral estimate.

To avoid this problem inherent in the tapering procedure, Thomson (1982) introduced the Multi-Taper Method of spectral analysis, in which the data are multiplied by not only one, but several leakage-resistant tapers. This results in several (S) tapered time series, $a_s(t)F(t),\ s = 1 \ldots S$, from one original data record $F(t)$. The tapers are orthogonal, so each of them samples the time series in a different manner while optimizing resistance to spectral leakage. The statistical information discarded by the first taper is partially recovered by the second taper, the information discarded by the first two tapers is partially retrieved by the third, and so on. Only a few low-order tapers may be employed, as the spectral leakage increases with increasing order and the high-order tapers allow an unacceptable level of leakage. Taking Fourier Transforms of each of the S tapered time series, S estimates of the spectrum, $Y_s(f)$, are produced. The tapers a_s are called "eigentapers" since their construction involves an eigendecomposition (see below). Similarly, the spectra $Y_s(f)$ are commonly called "eigenspectra". As will be shown below, the multi-taper spectral estimate is then formed as a weighted sum of the eigenspectra. Therefore, it is smoother than that obtained by using only one taper. The multi-taper spectrum has less variance than single-taper spectral estimates and at the same time is resistant to spectral leakage since the individual eigentapers are designed to be so.

A particularly useful family of orthogonal tapers with good leakage properties are the Discrete Prolate Spheroidal Sequences (DPSS), also known as Slepian tapers (Slepian, 1978; Thomson, 1982). The DPSS are obtained from the following $N \times N$ eigenproblem:

$$\mathbf{A}\ a_s \ = \ \lambda_s\ a_s \tag{3.1}$$

where a_s is a vector of length N containing the sth eigentaper and matrix

A has the form $A_{mn} = 2\Delta t W \sin(2\pi W(n-m)\Delta t)$ (Percival and Walden, 1993). The eigenvalue λ_s measures the resistance to spectral leakage of the corresponding eigentaper a_s. Eigentapers with $\lambda_s \sim 1$ can be used to construct spectral estimates that are resistant to spectral leakage. The "time-frequency bandwidth parameter" $p = NW$ defines a particular family of Slepian tapers as sequences of length N that have the largest possible concentration of the energy in the frequency interval $[-W, +W]$, where $W = pf_R$, and $f_R = (N\Delta t)^{-1}$ is the Rayleigh frequency of the time series (Δt is the sampling interval). This means that most of the energy in their spectral windows (that is, the Fourier Transform of the eigentapers themselves) is concentrated in the central lobe defined by the interval $[-W, +W]$, and hence the side lobes are small. This characteristic of the spectral window reflects the resistance to spectral leakage of the taper. According to Park et al. (1987), only the first $S = 2p-1$ tapers are usefully resistant to spectral leakage. Figures 3.1 and 3.2 provide an example of a family of Slepian tapers and their respective tapered time series, spectral windows and spectral estimates (modified from Percival and Walden, 1993).

The choice of the parameter p (and hence S) represents a trade-off between spectral resolution and the variance of the spectral estimate. If we make W large (which implies large p), the number of tapers with good leakage properties increases, hence we can make S large and thus reduce the variance of the spectrum. On the other hand, the resolution of the spectrum decreases as W increases. Therefore, if we make W too large we can inadvertently smear out fine features in the spectrum. In the context of climate studies of roughly century duration, the choice $p = NW = 2$ and hence $S = 3$ provides a good compromise between the resolution appropriate to resolve climatic signals, and the variance of the spectral estimate (Mann and Park, 1994, 1996).

For a given time series $F(t)$, we thus determine a set of S orthogonal Slepian data tapers $a_s(t)$ and their S associated tapered Fourier transforms or eigenspectra $Y_s(f), s = 1 \ldots S$, as:

$$Y_s(f) \;=\; \sum_{t=1}^{N} a_s(t)\; F(t)\; e^{i\,2\pi\, f\, t\, \Delta t} \tag{3.2}$$

where $F(t), t = 1 \ldots N$ is the time series, $a_s(t)$ is the sth member in a family of orthogonal Slepian tapers, with $s = 1 \ldots S$, and Δt is the sampling interval (monthly, seasonal, annual, etc). Only spectral fluctuations at frequencies greater than the Rayleigh frequency $f_R = (N\Delta t)^{-1}$ can be resolved. The frequency resolution of the spectrum is given by $2pf_R$. For example, considering a record of 100 years of monthly data (that is $N = 1200$ and $\Delta t = 1/12 = 0.083$ to obtain units in years) and using a bandwidth parameter $p = 2$, the frequency resolution of the spectrum is $2 \times 2/(1200 \times 0.083) = 0.04$ cycles/year,

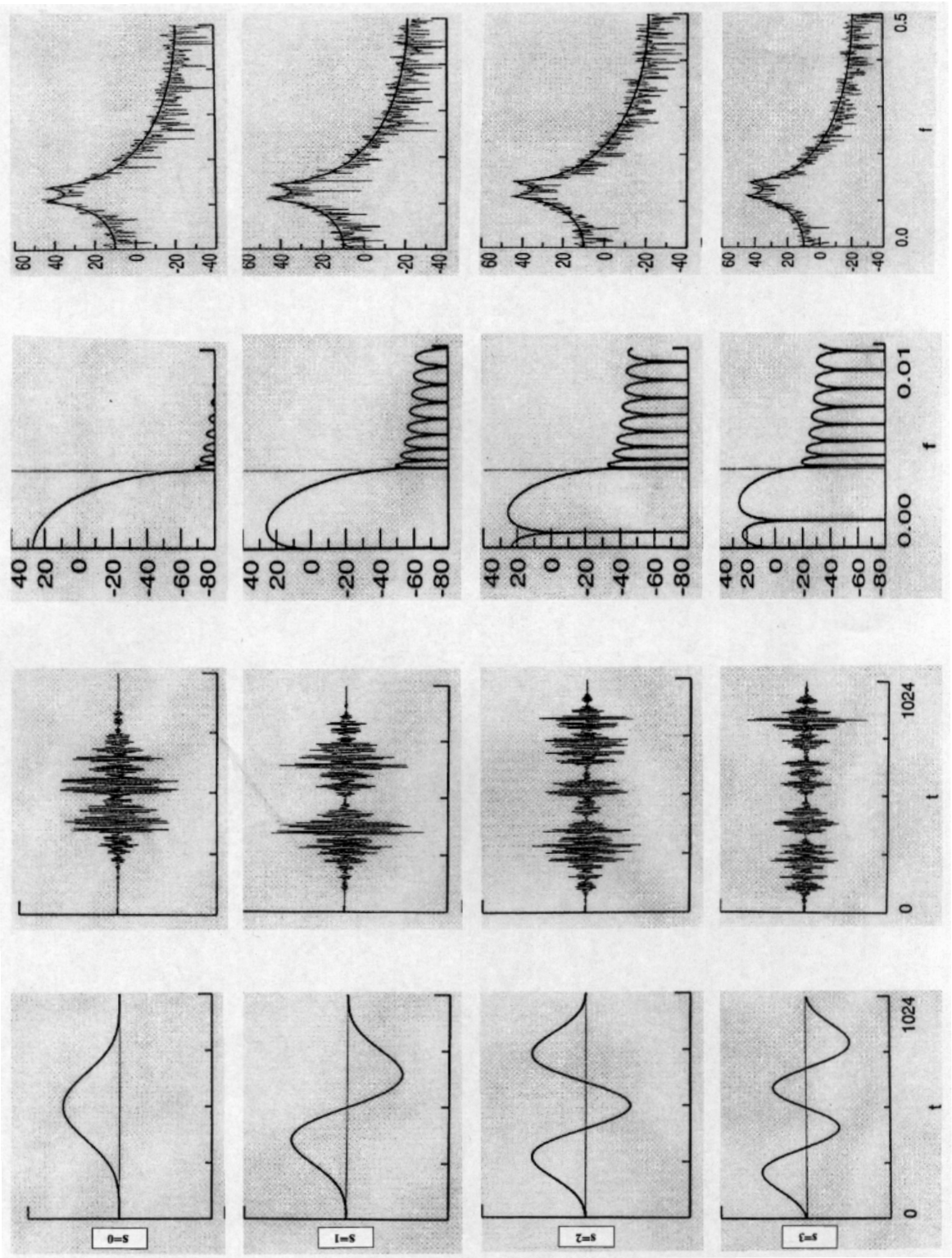

Figure 3.1: First column: Discrete Prolate Spheroidal Sequence (DPSS) data tapers $a_s(t)$ for $s = 0, 1, 2, 3$. Second column: The products $a_s(t)F(t)$ of the tapers and the time series $F(t)$. Third column: Spectral windows corresponding to the tapers $a_s(t)$. Fourth column: eigenspectra $Y_s(f)$ corresponding to the tapered time series $a_s(t)F(t)$. Adapted from Percival and Walden (1993).

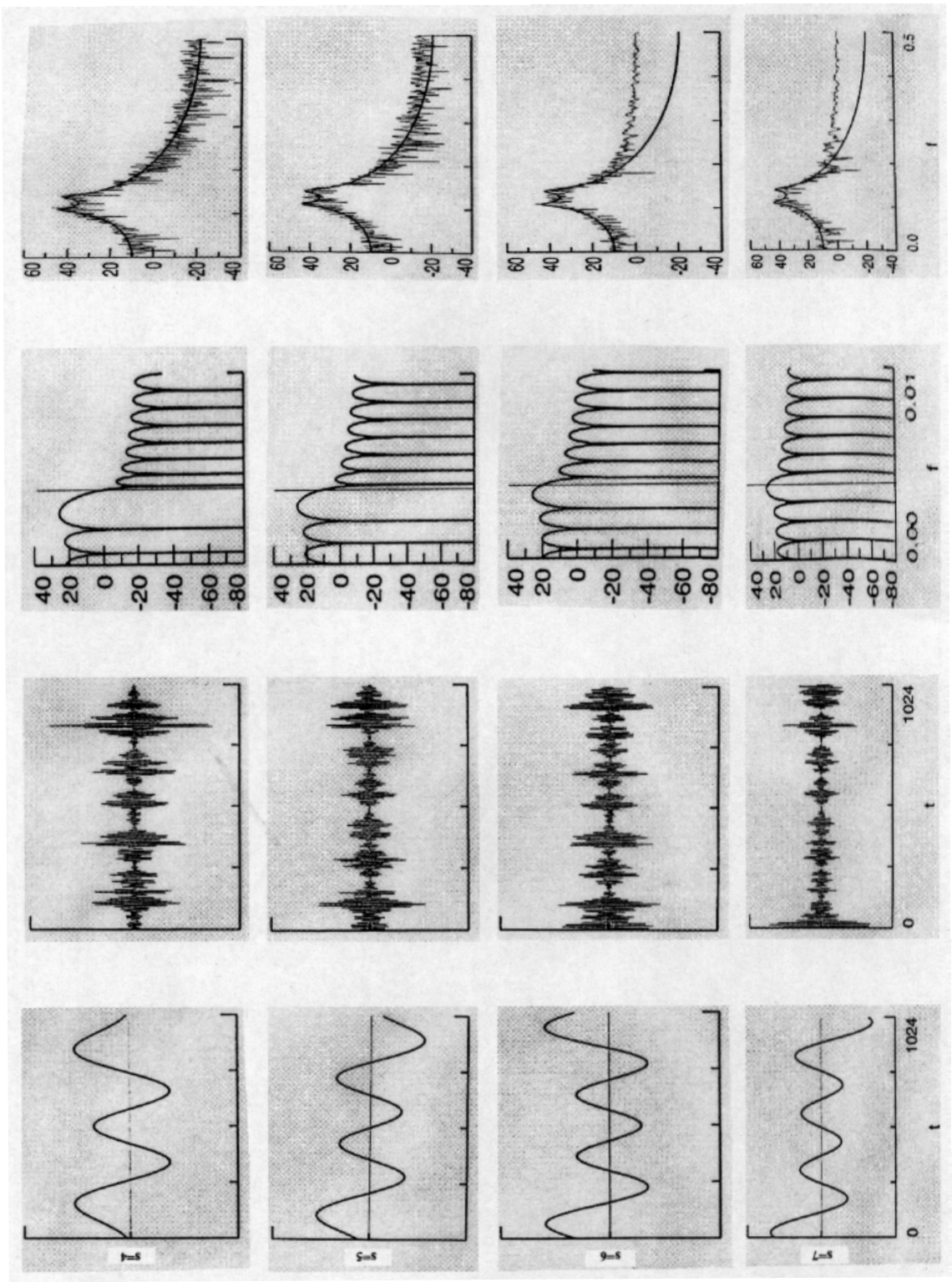

Figure 3.2: Same as in Figure 3.1 for taper orders $s = 4, 5, 6, 7$. Adapted from Percival and Walden (1993).

which allows to resolve decadal from interdecadal signals.

The multi-taper spectral estimate can provide a description of an irregular oscillatory signal centered at a any particular frequency f, since it can describe a variety of amplitude and phase modulations using a suitable linear combination of the S independent eigenspectra. For example, the independent eigenspectra can be combined through a weighted average as follows:

$$\overline{Y}(f) \;=\; \frac{\sum_{s=1}^{S} \lambda_s \, |Y_s(f)|^2}{\sum_{s=1}^{S} \lambda_s} \tag{3.3}$$

where the eigenvalues λ_s come from the decomposition in equation (3.1). The linear combination of eigenspectra given by $\overline{Y}(f)$ provides a power spectrum with optimal trade-off properties between spectral resolution and variance (Thomson, 1982; Park et al., 1987). Figure 3.3 shows examples of multi-taper spectral estimates $\overline{Y}(f)$ obtained as linear combinations of different numbers of eigenspectra (modified from Percival and Walden, 1993).

Further details on the tapering procedure and the MTM approach for spectral analysis can be found in Thomson (1982); Park et al. (1987); and Percival and Walden (1993). Some climate applications of the MTM technique are found in Kuo et al. (1990); Park and Maasch (1993); Mann and Park (1993); and Thomson (1995). A Toolkit to perform MTM spectral analysis is provided by the SSA-MTM Group at http://www.atmos.ucla.edu/tcd/ssa.

3.2 Singular Spectrum Analysis (SSA)

Singular Spectrum Analysis (SSA) is a variation of the classical EOF decomposition, but the application of the mathematics is substantially different. In the EOF analysis described in section 2.1, the field F to be studied consists of measurements obtained at a given time, that is, the coordinates of F represent "simultaneous" observations at different locations in space. Therefore, by solving the eigenproblem on the covariance matrix of F, we try to capture the dominant spatial patterns. The SSA expansion is an EOF expansion in which the field F contains values at the *same* location but at *different* time lags. The leading eigenvectors of the corresponding covariance matrix represent thus the leading *time patterns* of field F. SSA is a time series analysis technique, in the sense that a single signal (a time series) is analyzed, and it aims to identify recurrent patterns in time.

Let's start with a single time series $F(t)$, for $t = 1 \ldots N$, which may be

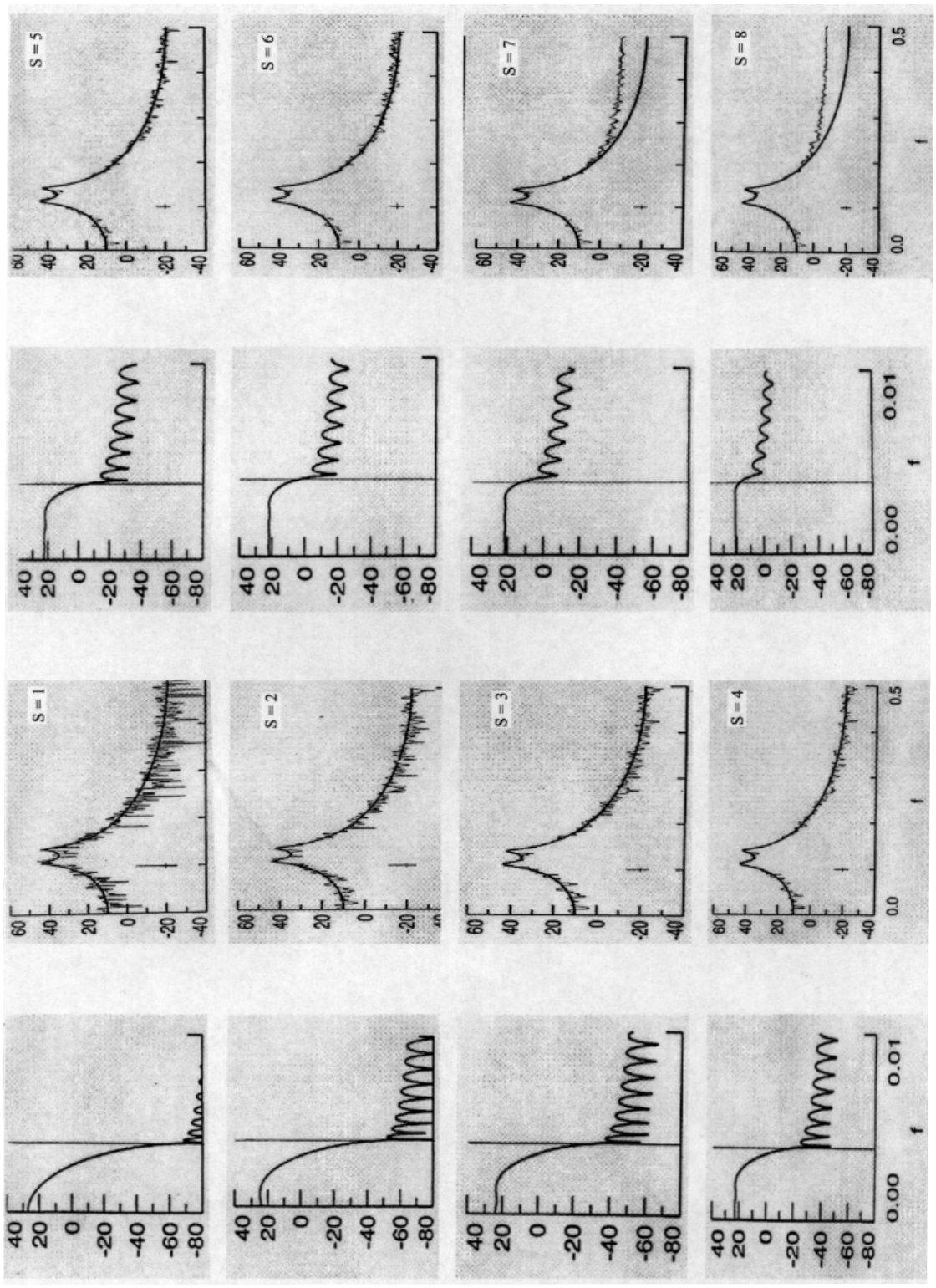

Figure 3.3: Multi-taper spectral estimates $\overline{Y}(f)$ formed by linear combinations of S eigenspectra (right column) and their respective spectral windows (left column), for $S = 1 \ldots 8$. The thin vertical lines on the spectral windows indicate the frequency W. Adapted from Percival and Walden (1993).

standarized using equation (2.1) (although this is not strictly necessary in the SSA method). We construct a data matrix $\mathbf{F}$ that contains the "lagged" time series $F(t+l\Delta)$, where $l = 0 \ldots L$ are the lags and Δ is the time increment (the "size" of the lag). Note that the lags start counting at $l = 0$, which means that the original (zero-lag) time series $F(t)$ is included in our definition $F(t + l\Delta)$. Some authors, however, do not include the zero-lag time series in matrix $\mathbf{F}$ but this difference is irrelevant to the results. Taking $\Delta = 1$ to make the notation simpler, data matrix $\mathbf{F}$ is:

$$\mathbf{F} = \begin{bmatrix} F(1) & F(2) & \ldots & F(N–L) \\ F(2) & F(3) & \ldots & F(N–L+1) \\ F(3) & F(4) & \ldots & F(N–L+2) \\ \ldots & \ldots & \ldots & \ldots \\ F(L+1) & F(L+2) & \ldots & F(N) \end{bmatrix} \tag{3.4}$$

Matrix $\mathbf{F}$ has dimension $L+1 \times N–L$. We then construct the "lagged" covariance matrix $\mathbf{R_{FF}}$ from data matrix $\mathbf{F}$ using equation (2.5):

$$\mathbf{R_{FF}} = \mathbf{F} * \mathbf{F}^{\dagger} \tag{3.5}$$

The lagged covariance matrix $\mathbf{R_{FF}}$ has dimension $L+1 \times L+1$ and contains covariances $\langle F(t+l)\ F(t+l)\rangle$ between the time series $F(t+l)$ at all possible combinations of lags $l = 0 \ldots L$:

$$\mathbf{R_{FF}} = \begin{bmatrix} \langle F(1)\ F(1)\rangle & \langle F(1)\ F(2)\rangle & \ldots & \langle F(1)\ F(L+1)\rangle \\ \langle F(2)\ F(1)\rangle & \langle F(2)\ F(2)\rangle & \ldots & \langle F(2)\ F(L+1)\rangle \\ \ldots & \ldots & \ldots & \ldots \\ \langle F(L+1)\ F(1)\rangle & \langle F(L+1)\ F(2)\rangle & \ldots & \langle F(L+1)\ F(L+1)\rangle \end{bmatrix} \tag{3.6}$$

The covariance matrix $\mathbf{R_{FF}}$ is called the Toeplitz Matrix, since is has constant diagonals. Its principal diagonal contains the variance of F, the second diagonal contains the lag $\pm$ 1 covariance of F, the third diagonal contains the lag $\pm$ 2 covariance of F, and so on.

As in traditional (spatial) EOF analysis, we now solve the eigenproblem on matrix $\mathbf{R_{FF}}$ using equation (2.8):

$$\mathbf{R_{FF}} * \mathbf{E} = \mathbf{E} * \mathbf{\Lambda} \tag{3.7}$$

As usual, $\mathbf{\Lambda}$ is the $L+1 \times L+1$ diagonal matrix containing the eigenvalues λ_k of $\mathbf{R_{FF}}$ as shown in matrix (2.9) and matrix $\mathbf{E}$ contains the eigenvectors of $\mathbf{R_{FF}}$. As in section 2.1.2, only the largest K eigenvalues are non zero, and so the effective dimension of $\mathbf{E}$ is $L+1 \times K$. In contrast with spatial EOF, the eigenvectors in matrix $\mathbf{E}$ are now *time patterns*, and are usually called Time-EOFs or T-EOFs. Each eigenvector E_l^k is a lagged sequence of length $L+1$, where $l = 0 \ldots L$ are the lags, and $k = 1 \ldots K$ are the modes:

$$\mathbf{E} = \begin{bmatrix} E_0^1 & E_0^2 & \ldots & E_0^K \\ E_1^1 & E_1^2 & \ldots & E_1^K \\ E_2^1 & E_2^2 & \ldots & E_2^K \\ \ldots & \ldots & \ldots & \ldots \\ E_L^1 & E_L^2 & \ldots & E_L^K \end{bmatrix} \tag{3.8}$$

The principal components $A^k(t)$ associated with these T-EOFs are called Time-PCs or T-PCs. They are calculated as usual by projecting the eigenvectors E_l^k onto the original time series $F(t+l)$ as in equations (2.11)- (2.12):

$$\mathbf{A} = \mathbf{E}^\dagger * \mathbf{F} \tag{3.9}$$

$$A^k(t) = \sum_{l=0}^{L} E_l^k \; F(t+l) \tag{3.10}$$

Matrix $\mathbf{A}$ is $K \times N-L$, so that each T-PC (contained in the rows of $\mathbf{A}$) has length $N-L$ as the lagged time series in $\mathbf{F}$. The T-PCs can be interpreted as moving averages of the original time series, the averages being weighted by the coordinates of the T-EOFs. The T-PCs are therefore filtered versions of the original time series $F(t+l)$. This has important spectral implications, as for example, that the sum of the spectra of the K T-PCs is equal to the spectrum of the original time series F. For further spectral characteristics of the T-PCs see Vautard et al. (1992).

In contrast with standard spectral analysis, in which the basis functions are given a priori as the sines and cosines of the Fourier expansion, in SSA they are determined from the data themselves to form an orthogonal basis that is optimal in the statistical sense. In SSA analysis, any oscillatory behavior present in the original time series stands out as a pair of nearly equal eigenvalues in matrix $\mathbf{\Lambda}$. Their associated T-EOFs and T-PCs have similar time scale of oscillation and are nearly in quadrature, that is, out of phase by approximately $\pi/2$. These oscillatory modes may be detected as pairs of con-

secutive similar eigenvalues (λ_k, λ_{k+1}), consecutive T-EOFs (E_l^k, E_l^{k+1}) and T-PCs ($A^k(t), A^{k+1}(t)$). Because of this property, the SSA method is particularly helpful in isolating anharmonic oscillations with fluctuating amplitudes from noisy data. Several objective criteria have been developed by Vautard et al. (1992) in order to extract these oscillatory pairs. Furthermore, the associated T-PCs provide information on phase and amplitude. The quadrature relationship allows us to represent the state of the oscillation as a complex number from which the phase $\varphi(t)$ and the amplitude $\rho(t)$ can be determined from the knowledge of the T-PCs:

$$Z^k(t) \;=\; A^k(t) + i\; A^{k+1}(t) \;=\; \rho(t)\; e^{i\;\varphi(t)} \tag{3.11}$$

The portion of the variability in the original time series F that is associated to a given oscillation captured by a pair of modes can thus be isolated by restricting the SSA expansion to the T-EOFs and T-PCs corresponding to that pair of modes. That is, we can reconstruct the original time series using only the two SSA components of interest. The Reconstructed Components (RCs) which carry the contributions of the pair of modes k and $k+1$ to the variance of the original data can be computed as a special case of equation (2.14):

$$RC_l^{k,k+1}(t) \;=\; E_l^k\; A^k(t) \;+\; E_l^{k+1}\; A^{k+1}(t) \tag{3.12}$$

The RCs are additive and their complete sum (for all modes $k = 1 \ldots K$) gives back the original time series F.

The choice of the "embedding dimension" $m^* = L * \Delta$, where L is the number of lags and Δ is the size of the lag, is crucial in the SSA method. The measure m^* is equivalent to the width of a moving window passed through the time series. It must be chosen according to the timescale under study, but its value is restricted by the temporal dimension of the data (that is, the length of the time series, N). A rule of thumb is that SSA can usefully analyze only those quasi-oscillatory structures with periods in the range $\{m^*/5, m^*\}$, where $m^* = N/3$ (Vautard et al., 1992). Consequently, there are rather severe restrictions on the range of frequency bands over which the temporal structure of a time series can be reconstructed. For instance, to recover an interdecadal signal (of approximately 20 years period) in a time series of 100 years of annual data ($N = 100$), we may choose $m^* = 30$ years, according to the rule given above. However, this window width does not allow a reliable decomposition of oscillatory signals with dominant periods of less than 6 years ($m^*/5$). Hence, such a value of m^* is not well suited for the study of the ENSO signal, for

example. As such, the choice of the measure m^* has a subjective component, since it also depends on the timescale of interest of the analysis, while being restricted by the rule given above. We will extend the discussion on the characteristics of the measure m^* at the end of section 4.1. A multivariate version of the SSA, called Multichannel SSA or MSSA, is described in section 4.4. For further details on the SSA method and applications see Vautard et al. (1992); Keppenne and Ghil (1992); Dettinger et al. (1995); Allen and Smith (1996); Emery and Thomson (1998); Vautard (1999); and von Storch and Zwiers (1999). A Toolkit to perform SSA is provided by the SSA-MTM Group at http://www.atmos.ucla.edu/tcd/ssa.

3.3 Wavelet Analysis (WA)

The Wavelet Analysis (WA) is a time series analysis method that has increasingly been applied in geophysics during the last couple of decades. It is becoming a common tool for analyzing temporal variations of power within a time series. By transforming a time series from the time space into the time-frequency space, WA is able to determine both the dominant timescales of variability and how they vary with time. WA has several attractive advantages over the traditional Fourier Analysis, especially when dealing with time series with time-varying amplitudes. In contrast to the Fourier Transform, that generates values of amplitude and phase averaged over the entire time series for each frequency component or harmonic, the Wavelet Transform provides a localized, "instantaneous" estimate of the amplitude and phase for each spectral component of the series. This gives WA an advantage in the analysis of non-stationary data in which the amplitude and phase of the harmonic components may change rapidly in time or space. While the Fourier Transform of the non-stationary time series would smear out any detailed information on the changing features, the WA keeps track of the evolution of the signal characteristics throughout the time series.

WA is based on the convolution of a time series $F(t)$ with a set of functions $G_{ab}(t)$ with parameters a and b, derived from a "mother wavelet" or "analyzing wavelet" $G(t)$, where:

$$G_{ab}(t) \;=\; \frac{1}{a^{1/2}}\, G\left(\frac{t-b}{a}\right) \tag{3.13}$$

The real scalar b is called the "translation" parameter and corresponds to the central point of the wavelet in the time series. The real and positive scalar a is the "scale dilation" parameter and determines the width of the wavelet.

The factor $1/a^{1/2}$ normalizes the wavelets so that they have unit energy and hence are comparable for all scales a. Any set of functions of this form which satisfy the conditions outlined below are called "wavelets". The convolution of the time series $F(t)$, $t = 1 \ldots N$, with the set of wavelets $G_{ab}(t)$ for different parameters a and b, defines the Wavelet Transform (WT):

$$T(b,a) \;=\; \frac{1}{a^{1/2}} \sum_{t=1}^{N} G^{*}\left(\frac{t-b}{a}\right)\; F(t) \tag{3.14}$$

where the asterisk denotes complex conjugation and parameters a and b are allowed to vary continuously. The translation parameter b corresponds to time in the case of a time series (or position in the case of spatial series). The dilation parameter a corresponds to scale length or temporal period (inverse of frequency). By varying the scale a and translating along the localized time b we can construct a two-dimensional picture showing the amplitude of any feature in the time series as a function of the scale and also how this amplitude varies with time. Thus, WA expands a one-dimensional time series into a two-dimensional (a, b)-parameter space, with time b and scale a as new independent variables. The Fourier Transform maps a one-dimensional series into a one-dimensional spectrum (that is, transforms from time to frequency domain). The Wavelet Transform maps a one-dimensional series into a two-dimensional image that displays the evolution of the relative amplitude of the different scales (frequencies) with time (that is, transforms from time to time-frequency domain). The WT of equation (3.14) can be thought of as a sort of mathematical microscope, with magnification $1/a$, position b and optics given by the choice of the specific wavelet $G(t)$ (Shen et al., 1994). Whereas Fourier Analysis provides an average amplitude over the entire time series, WA yields a measure of the localized amplitudes a as the wavelet moves through the time series with increasing values of b. Although wavelets have a definite scale, they typically do not bear any resemblance to the sines and cosines of the Fourier decomposition. Nevertheless, a correspondence between wavelength and scale a can sometimes be found.

According to Meyers et al. (1993), the function $G(t)$ must satisfy the following properties in order to be a mother wavelet:

1. Its amplitude $|G(t)|$ must tend rapidly to zero as time $t \to \infty$. This requirement is responsible for the local nature of wavelet analysis, since the transformed values $T(b, a)$ are generated only by the signal inside the "cone of influence" (COI) centered at $t = b$. In practice, the radius of the COI is the point $t = r_c$ beyond which the wavelets $G_{ab}(t)$ no longer have significant values (their values are close to zero). Usually, r_c

is proportional to the scale a, which results in the cone-like structure of the WT.

2. Wavelet $G(t)$ must have zero mean (known as the "admissibility condition"). This ensures that the WT can be inverted, that is, the original signal $F(t)$ can be recovered from the wavelet coefficients through the inverse transform:

$$F(t) = \frac{1}{C} \sum_a \sum_b \frac{T(b,a)\, G_{ab}}{a^2} \tag{3.15}$$

where

$$\frac{1}{C} = \sum_f \frac{\left|\hat{G}(f)\right|^2}{f} \tag{3.16}$$

Here $\hat{G}(f)$ is the Fourier Transform of $G(t)$. For $1/C$ to remain finite, it must be $\hat{G}(0) = 0$.

3. Wavelets are often regular functions, so that $\hat{G}(f < 0) = 0$. This eliminates the confusion between measurements at $f > 0$ and $f < 0$, which simplifies the interpretation of the transform. Wavelets are described in terms of positive frequencies only ("progressive" wavelets).

4. Higher-order moments (variance, skewness, etc) should vanish to allow the investigation of higher-order moments in the data. This requirement can be relaxed, according to the application.

The appropriate choice of the form of the wavelet $G(t)$ depends on the goal of the study. It is clear that the description of a true physical signal should not depend on the choice of the wavelets. However, the best results are obtained when using wavelets that bear some resemblance in form to the signal. Therefore, if we know the characteristics of the signal being sought, we should choose a wavelet that has a similar pattern. Then, large values of the amplitude of the transform $T(b,a)$ will indicate where the time series $F(t)$ has the desired form (a form similar to the wavelet). One of the most commonly used wavelets in geophysics is the Morlet Wavelet:

$$G(t) = e^{-t^2/2}\, e^{ict} \tag{3.17}$$

It consists of a plane wave e^{ict} of frequency $c = f$ (or wavenumber $c = k$ in the spatial domain), which is modulated in time by a Gaussian envelope of unit width $e^{-t^2/2}$. A variation of this wavelet, applicable to a signal with two dominant frequencies c_1 and c_2 is:

$$G(t) = e^{-t^2/2}\, e^{ic_1 t}\, e^{ic_2 t} \tag{3.18}$$

A wavelet applicable to short segments of data with linearly increasing frequency is:

$$G(t) = e^{-t^2/2}\, e^{ict}\, e^{ikt^2/2} \tag{3.19}$$

Other types of wavelets include the "Mexican Hat", which is the second derivative of the Gaussian function, the simpler Haar Wavelet, which is based on a box function, and the series of Daubechies wavelets of different orders (see Daubechies, 1992; Torrence and Compo, 1998, for further descriptions).

A characteristic of the wavelets is that they can be stretched and translated with flexible resolution in both frequency and time. The flexible time-frequency windows are adaptive to the entire time-frequency domain. The latter narrows when focusing on high-frequency signals and broadens when searching for low-frequency signals. This "zoom-in" property is a unique characteristic of the wavelets, that allows for the localization of very short-lived, high-frequency signals in time, such as abrupt changes, while still resolving the low-frequency variability with reasonable accuracy.

There are some considerations about the choice of continuous or discrete values of parameters a and b. For example, a continuous wavelet transform (using "continuous" wavelets) can yield redundant information since small changes in a or b are often insignificant. It is more efficient to choose discrete values of a and b such that the functions $G_{ab}(t)$ constitute an orthogonal basis (using discrete or "orthogonal" wavelets). The transform at any value of a and b can be found a posteriori through a suitable interpolation.

There exist several methods for implementing a Wavelet Analysis. The simplest algorithm is the direct numerical integration of a discretisized form of equation (3.14). Knowing $F(t)$ and $G(t)$, we can compute the WT at discrete points in the time-frequency domain. The problem with this technique is

that it is extremely time consuming. An alternative method is to use the convolution theorem to obtain the WT in the frequency domain, using the Fourier Transforms of $G(t)$ and $F(t)$ as:

$$T(b,a) \;=\; \frac{1}{a^{1/2}} \sum_{f} \hat{G}^{*}(af)\; \hat{F}(f)\; e^{ibf} \tag{3.20}$$

where $\hat{G}(af)$ and $\hat{F}(f)$ are the Fourier transforms of $G(t/a)$ and $F(t)$, respectively, and the asterisk indicates complex conjugation. We can now compute the inverse Fourier transform of this expression to obtain the Wavelet Transform $T(b,a)$ in the time-frequency domain (parameters a, b domain). To use this technique, $\hat{G}(f)$ must be known analytically and the time series $F(t)$ must be pre-processed to avoid errors at the end points ("ringing") derived from the Fourier algorithms. For example, if $F(t)$ is aperiodic, equation (3.20) will introduce an artificial periodicity in the WT. Different methods for dealing with this problem are discussed in detail in Emery and Thomson (1998) and Meyers et al. (1993). The most advisable approach is to taper or buffer the original time series $F(t)$ with added data points that smoothly tend to zero at the beginning and end of the series. The region of the WT corresponding to these artificial data points is afterwards discarded from the analysis. If this buffering is not performed, the WT will show severe distortions near the ends of the series.

A practical procedure to perform a Wavelet Analysis can be summarized as follows. To reduce the discussed end problems, we should extend the time series $F(t)$ by adding a trigonometric taper of the form $1 - \sin\varphi$, as a "tail" that goes to zero at the beginning and end of the series (see Meyers et al., 1993; Emery and Thomson, 1998, for details). The final number of data points in the series should be a power of 2, that is $N = 2^m$ where m is an integer, to facilitate the Fast Fourier Transformation. Then, we remove the record mean of the tapered time series and compute its Fourier Transform $\hat{F}(f)$. We should also compute the Fourier Transform of the wavelet $G(t)$ at specific scales a, and obtain $\hat{G}(af)$. Then, we calculate the Wavelet Transform as in equation (3.20), by convolving the product $\hat{G}^{*}(af)\hat{F}(f)$ in Fourier space. Finally, we take the inverse Fourier Transform of the result to obtain $T(b,a)$ as a function of time b and scale a. Since the wavelet function $G(t)$ is complex, the resulting Wavelet Transform $T(b,a)$ is also complex. As such, it can be written as the sum of a real part $\Re\{T(b,a)\}$ and an imaginary part $\Im\{T(b,a)\}$, or as a combination of an amplitude $|T(b,a)|$ and a phase arctan ($\Im\{T(b,a)\}$ / $\Re\{T(b,a)\}$). In addition, we can define the "wavelet power spectrum" as $|T(b,a)|^2$. We can thus present the WT by plotting the wavelet power spectrum as a function

of time b in the linear x-axis and scale a in the logarithmic y-axis. To make it easier to compare different wavelet power spectra, we can normalize their values as $|T(b,a)|^2/\sigma^2$, where σ is the variance of $F(t)$. The normalization $1/\sigma$ gives a measure of the power relative to white noise, since σ is the expectation value of the WT of a white noise process for all values of a and b. For extended descriptions of Wavelet Analysis and applications see Meyers et al. (1993); Gamage and Blumen (1993); Liu (1994); Weng and Lau (1994); Lau and Weng (1995); Gu and Philander (1997); Wang and Wang (1996); Torrence and Compo (1998); and Emery and Thomson (1998). Further information on WA can be also found in http://www.amara.com/current/wavelet.html.

Chapter 4

Patterns in Space and Time: Signal Propagation

4.1 Extended Empirical Orthogonal Functions (EEOF)

Conventional EOF analysis, as we have demonstrated in section 2.1, is very successful in compressing the complicated variability of the original data set into the fewest possible number of modes while retaining most of the total variance. However, it can only represent variations in time that are in phase or out of phase along a data array, that is, standing oscillations. It cannot represent features that have variable phase relationships, such as propagating waves or moving structures. This is a strong limitation since many geophysical phenomena result from the interaction between traveling waves of different spatial scales and different frequencies.

In order to detect propagating phenomena in multivariate data sets, EOF analysis has been modified in different ways. One of the most popular variations is the Extended EOF analysis (EEOF), introduced in a geophysical context by Weare and Nasstrom (1982). In EEOF, a single EOF can represent propagating features since the technique determines the eigenvectors of the lagged covariance matrix which is derived from the observations sampled at a limited number of successive time steps or lags. Since this approach becomes excessively expensive in computer time as the number of lags is increased, it is only useful to resolve propagating phenomena over a limited time span. In a way, we could regard the EEOF method as a multivariate version of the SSA procedure described in section 3.2.

The EEOF method identifies the dominant spatial and temporal structure of lagged sequences of covariances. Such a procedure is able to capture time-evolving patterns in the data since phase information is retained in the

decomposition. The approach is useful to recover oscillatory patterns existing in the data, but requires some a priori knowledge of the dominant timescales in the data or a particular interest on a specific timescale, since it needs a subjective choice of the number of lags used and their size, as is explained below.

In traditional EOF analysis, the data matrix $\mathbf{F}$ contains field values measured at different spatial locations simultaneously (at a given time). In EEOF analysis, however, the data matrix $\mathbf{F}$ contains field values measured at different locations and at different time lags. EEOF analysis aims to investigate the spatial and temporal covariability of a field with itself for a number of locations and for a number of time steps into the future or into the past.

The procedure to construct the data matrix $\mathbf{F}$ in EEOF bears some similarity with that in SSA (section 3.2) in that it uses lagged versions of field $F(t)$, and with that in Combined EOF (section 2.2) in that it concatenates two or more fields in one data matrix. However, instead of combining different fields, here we combine field $F_m(t)$ with "lagged" versions of itself at different time lags $l = 1 \ldots L$. That is, we combine $F_m(t)$ with $F_m(t+1*\Delta)$, $F_m(t+2*\Delta)$, ..., $F_m(t+L*\Delta)$, where Δ is the time increment (the "size" of the lag). For simplicity, we consider $\Delta = 1$ in the following. Data matrix $\mathbf{F}$ is thus constructed by concatenating the original field $F_m(t)$ and the lagged fields $F_m(t+l)$ for all locations $m = 1 \ldots M$, time steps $t = 1 \ldots N-L$, and lags $l = 1 \ldots L$ as follows:

$$\mathbf{F} = \begin{bmatrix} F_1(1) & F_1(2) & \ldots & F_1(N-L) \\ F_2(1) & F_2(2) & \ldots & F_2(N-L) \\ \ldots & \ldots & \ldots & \ldots \\ F_M(1) & F_M(2) & \ldots & F_M(N-L) \\ F_1(2) & F_1(3) & \ldots & F_1(N-L+1) \\ F_2(2) & F_2(3) & \ldots & F_2(N-L+1) \\ \ldots & \ldots & \ldots & \ldots \\ F_M(2) & F_M(3) & \ldots & F_M(N-L+1) \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ F_1(L+1) & F_1(L+2) & \ldots & F_1(N) \\ F_2(L+1) & F_2(L+2) & \ldots & F_2(N) \\ \ldots & \ldots & \ldots & \ldots \\ F_M(L+1) & F_M(L+2) & \ldots & F_M(N) \end{bmatrix} \tag{4.1}$$

Data matrix $\mathbf{F}$ has dimension $(L+1)*M \times N-L$. Some authors do not include the zero-lag (original) version of F_m in this matrix, which results in a matrix $\mathbf{F}$ of dimension $L*M \times N-L$. The "lagged" covariance matrix $\mathbf{R_{FF}}$ is then

computed in the usual way from matrix $\mathbf{F}$, using equation (2.5):

$$\mathbf{R_{FF}} = \mathbf{F} * \mathbf{F}^{\dagger} \tag{4.2}$$

Matrix $\mathbf{R_{FF}}$ is now a huge $(L{+}1){*}M \times (L{+}1){*}M$ matrix whose elements are the covariances between all possible combinations of field F at different locations m *and* at different lags l. That is, the elements of the lagged covariance matrix are the simultaneous covariances $\langle F_i(t), F_j(t)\rangle$ (as in traditional EOF) *and also* the lagged covariances $\langle F_i(t{+}l), F_j(t{+}l)\rangle$, where i, j span all locations $1 \ldots M$, and l spans all the lags $1 \ldots L$. It is easy to see that the size of the covariance matrix increases dramatically with the number of lags used. Therefore, the application of EEOF analysis needs to be limited to a small number of lags in order to make the computations possible.

We now solve the eigenproblem on the lagged covariance matrix $\mathbf{R_{FF}}$ as usual using equation (2.8):

$$\mathbf{R_{FF}} * \mathbf{E} = \mathbf{E} * \mathbf{\Lambda} \tag{4.3}$$

where all matrices have dimension $(L{+}1){*}M \times (L{+}1){*}M$. Eigenvalues λ_k in matrix $\mathbf{\Lambda}$ are sorted in decreasing order and are associated with eigenvectors E^k_{ml} in the columns of matrix $\mathbf{E}$, where k indicates the mode. As in section 2.1, only the first $K \leq min\{(L{+}1){*}M, N{-}L\}$ eigenvalues are non zero and thus the effective dimension of matrix $\mathbf{\Lambda}$ is $K \times K$ and that of matrix $\mathbf{E}$ is $(L{+}1){*}M \times K$. The length of each eigenvector E^k_{lm} is now $(L{+}1){*}M$ (number of lags $\times$ number of locations), hence the combined subindex lm. Each eigenvector E^k_{lm} consists of a sequence of $L+1$ EOF patterns for lags $l = 0 \ldots L$, one after the other, as follows:

$$\mathbf{E} = \begin{bmatrix} E_1^1 & E_1^2 & \dots & E_1^K \\ E_2^1 & E_2^2 & \dots & E_2^K \\ \dots & \dots & \dots & \dots \\ E_M^1 & E_M^2 & \dots & E_M^K \\ E_{11}^1 & E_{11}^2 & \dots & E_{11}^K \\ E_{12}^1 & E_{12}^2 & \dots & E_{12}^K \\ \dots & \dots & \dots & \dots \\ E_{1M}^1 & E_{1M}^2 & \dots & E_{1M}^K \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ E_{L1}^1 & E_{L1}^2 & \dots & E_{L1}^K \\ E_{L2}^1 & E_{L2}^2 & \dots & E_{L2}^K \\ \dots & \dots & \dots & \dots \\ E_{LM}^1 & E_{LM}^2 & \dots & E_{LM}^K \end{bmatrix} \tag{4.4}$$

That means that an EOF pattern in EEOF analysis is no longer a single map of dimension M as in traditional EOF, but it is instead a *sequence* of $L+1$ lagged maps (each of dimension M). A sequential observation of these consecutive spatial patterns provides information about possible propagating features associated with each mode.

Principal components $A_l^k(t)$ for mode k and lag l are obtained by projecting the EOF pattern E_{lm}^k for a given mode k and lag l onto the original time series in $\mathbf{F}$ lagged at the same lag l, that is $F_m(t+l)$ (the portion of matrix (4.1) for lag l). Using equation (2.11):

$$A_l^k(t) \;=\; \sum_{m=1}^{M} E_{lm}^k \; F_m(t+l) \tag{4.5}$$

Each principal component A_l^k has length $N-L$, which is the length of any of the original time series $F_m(t+l)$. Therefore, mode k of the EEOF decomposition consists of a sequence of $L+1$ EOFs E_{ml}^k, a sequence of principal components A_l^k and an eigenvalue proportional to the variance accounted for by the mode (as in equation (2.13)).

As discussed in section 3.2, the choice of the "embedding dimension" m^*, that is, the number of lags L and their time increment Δ, is crucial and rather subjective. It depends on the number of data points in the time series and the frequency range under study (see section 3.2). If the embedding dimension is too small compared to the period of oscillation of interest, the sequence

of spatial patterns spans only a small portion of a complete oscillation. Any propagation of features relevant to the mode of interest results too "slow" to be clearly distinguished from one map to the next and the information obtained from such an analysis is rather incomplete. On the other extreme, if the lags are too long compared to the period of oscillation of interest, the sequence of spatial patterns may "miss" the signal under study, since its timescale may be smaller than the interval between two maps. The choice of the number of lags and their size is closely related to the concept of spectral resolution.

As an illustrative example of this choice, let's imagine that we want to isolate the ENSO phenomenon using lagged time series of monthly sea surface temperature in the Equatorial Pacific. We know a priori that the timescale of the ENSO signal is about 3-5 years, so this is the period of interest of our study. Then, if we choose to perform an EEOF analysis with, for instance, 3 lags of 1 month each ($L=3$, $\Delta=1$ month), the sequence of spatial EOF patterns will cover 3 months, which only spans a very small fraction of an average ENSO oscillation. Any propagation of structures related to ENSO becomes difficult to assess in such a small fraction of a cycle. On the other extreme, we may choose to use 3 lags of 5 years each ($L = 3$, $\Delta = 5 * 12$ months), in which case the sequence of spatial patterns will span 15 years and cannot detect the 3-5-yr oscillations that will be hidden by the coarse resolution of the sequence. A reasonable choice of L and Δ for this study would be for instance to use 8 lags of 6 months each ($L=8$, $\Delta=6$ months). This provides enough temporal resolution between maps to detect propagation of the ENSO signal and a view of one complete average cycle: $m^* = L*\Delta=8*6=48$ months (4 years).

The EEOF method is mathematically equivalent to the Multichannel Singular Spectrum Analysis (MSSA) method discussed in section 4.4. In practice, however, the two techniques slightly differ in their domain of application. In general MSSA deals with more temporal than spatial degrees of freedom (that is, more lags than locations or $L \gg M$), whereas in EEOF the data matrix contains only a few lags and a large number of spatial locations (that is, $L \ll M$). See section 4.4 for details on MSSA. Additional reading and applications of the EEOF method can be found in Weare and Nasstrom (1982); Lau and Chan (1986); Graham et al. (1987); Preisendorfer (1988); White and Peterson (1996); Allen and Robertson (1996); Tourre and White (1997); and Peterson and White (1998).

4.2 Frequency-Domain Empirical Orthogonal Function (FDEOF)

A complex generalization of EOF analysis for the study of propagating phenomena is the Frequency-Domain EOF (FDEOF), also known as Complex

Harmonic EOF, developed in a meteorological context by Wallace and Dickinson (1972). The technique involves computing complex eigenvectors from spectral matrices and is the most general of the available methods for analyzing propagating structures. However, due to the extremely general and all-embracing nature of the approach, the application of FDEOF analysis on climate problems becomes troublesome. For example, if the energy of an EOF mode is spread over a wide range of frequencies, that is, when the climatic fluctuations occur at irregular intervals, then many maps (one for each spectral estimate) are needed to document one single phenomenon. Therefore, the application of FDEOF in climate research is limited and the use of alternative methods, such as the Complex EOF analysis described in the next section, may be preferred. The FDEOF technique is briefly outlined here for completeness. Further details can be found in Wallace and Dickinson (1972); Brillinger (1981); and Preisendorfer (1988).

Time series $F_m(t)$ for locations $m = 1 \ldots M$ are first transformed from the time domain into the frequency domain using Discrete Fourier Transform (DFT). The spectral estimates of each $F_m(t)$ are called $Y_m(f)$:

$$Y_m(f) \;=\; \sum_{t=1}^{N} F_m(t)\; e^{i\,2\pi\,f\,t\,\Delta} \tag{4.6}$$

where f are the frequencies and Δ is the sampling interval. A matrix $\mathbf{Y}$ is thus constructed with the complex values of the spectral estimates $Y_m(f)$ of the time series, for all the calculated values of the frequency $f = f_1 \ldots f_n$ where f_n is the Nyquist frequency (that is, the highest detectable frequency determined by the sampling interval Δ between the data points). Each row of complex matrix $\mathbf{Y}$ contains the values of the spectral estimate of F_m for all frequencies at a given location m. Each column contains the values of the spectral estimate for all locations at a given frequency f:

$$\mathbf{Y} = \begin{bmatrix} Y_1(f_1) & Y_1(f_2) & \ldots & Y_1(f_n) \\ Y_2(f_1) & Y_2(f_2) & \ldots & Y_2(f_n) \\ \ldots & \ldots & \ldots & \ldots \\ Y_M(f_1) & Y_M(f_2) & \ldots & Y_M(f_n) \end{bmatrix} \tag{4.7}$$

An eigenvalue decomposition is now performed on matrix $\mathbf{Y}$. This may be done using any of the two approaches described in section 2.1, that is, constructing a covariance matrix and decomposing it into eigenvectors and eigenvalues, or using the singular value decomposition. Here we choose to use the

singular value decomposition approach (section 2.1.4). Using equation (2.19) on matrix $\mathbf{Y}$ we obtain:

$$\mathbf{Y} = \mathbf{U} * \mathbf{\Gamma} * \mathbf{V}^{\dagger} \tag{4.8}$$

$$Y_m(f) = \sum_{k=1}^{K} U_m^k \, \gamma_k \, V^{\dagger k}(f) \tag{4.9}$$

Singular vectors U_m^k are the complex spatial patterns (EOFs), which now contain amplitude and phase information. Singular vectors $V^{\dagger k}(f)$ are the analogous of the principal components, but they are now in the frequency domain. That is, they are combinations of pure harmonic components that describe the relatively smooth time evolution of the kth spatial pattern. The singular values, as usual, determine the fraction of the variance explained by the mode. Since phase information is maintained in this procedure, both standing and travelling oscillatory signals in the data can be described. As we mentioned above, however, the FDEOF technique has important limitations, especially when the data presents irregular or modulated oscillations. Some of the limitations are largely overcome by the use of the Complex or Hilbert EOF method described next.

4.3 Complex or Hilbert Empirical Orthogonal Functions (CEOF)

An attractive alternative to the traditional FDEOF is the Complex EOF analysis (CEOF), also called Hilbert EOF analysis. This approach has gained popularity since it avoids the problems associated with conventional FDEOF and it is of simpler application. Complex time series are "artificially" formed using the original time series and their Hilbert transforms. Then, complex eigenvectors are determined from the cross-covariance or cross-correlation matrix derived from the complex time series. The CEOF analysis allows for an efficient detection of propagating features, especially when their variance is spread over a number of frequencies, as is very often the case in climate data.

Let's consider the original field $\phi_m(t)$ as in section 2.1.1 (before mean removal and normalization), where $m = 1 \ldots M$ are the locations and $t = 1 \ldots N$ are the time steps. The M time series are then "made complex" by adding their Hilbert transforms $\hat{\phi}_m(t)$ as an artificial imaginary component, namely:

$$\Phi_m(t) \;=\; \phi_m(t) + i\;\hat{\phi}_m(t) \tag{4.10}$$

Let the Fourier representation of $\phi_m(t)$ be:

$$\phi_m(t) \;=\; \sum_{\omega}\; a_m(\omega)\;\cos\omega t + b_m(\omega)\;\sin\omega t \tag{4.11}$$

Then, the Hilbert transform $\hat{\phi}_m(t)$ of field $\phi_m(t)$ is:

$$\hat{\phi}_m(t) \;=\; \sum_{\omega}\; b_m(\omega)\;\sin\omega t - a_m(\omega)\;\cos\omega t \tag{4.12}$$

The Hilbert transform $\hat{\phi}_m(t)$ provides information about the rate of change of the time series at each time t. It is also called "quadrature function" since it is easy to see from the definitions that $\hat{\phi}_m(t)$ represents $\phi_m(t)$ phase-shifted by $\pi/2$. The Hilbert transform thus represents a filtering operation upon $\phi_m(t)$ in which the amplitude of each spectral component is unchanged but the phase of each spectral component is advanced by $\pi/2$. For further details on the Hilbert transform see, for example, Horel (1984).

The "complexified" observations $\Phi_m(t)$ are then transformed to standarized time series $F_m(t)$ by removing the temporal mean and dividing by the standard deviation at each spatial position m as in equation (2.1) (note that the standarization should be done for the real and the imaginary parts separately).

We now organize the complex observations $F_m(t)$ into a $M \times N$ data matrix $\mathbf{F}$ as in equation (2.4):

$$\mathbf{F} = \begin{bmatrix} F_1(1) & F_1(2) & \ldots & F_1(N) \\ F_2(1) & F_2(2) & \ldots & F_2(N) \\ \ldots & \ldots & \ldots & \ldots \\ F_M(1) & F_M(2) & \ldots & F_M(N) \end{bmatrix} \tag{4.13}$$

We then compute the complex covariance (strictly correlation) matrix $\mathbf{R_{FF}}$ of the complex field $F_m(t)$ in the usual way:

$$\mathbf{R_{FF}} = \mathbf{F} * \mathbf{F}^* \tag{4.14}$$

where the asterisk denotes complex conjugation (the equivalent of the transpose for complex matrices). The complex matrix $\mathbf{R_{FF}}$ has dimension $M \times M$ and is a complex version of matrix (2.6). We thus perform a decomposition of the complex covariance matrix $\mathbf{R_{FF}}$ using equation (2.8):

$$\mathbf{R_{FF}} * \mathbf{E} = \mathbf{E} * \mathbf{\Lambda} \tag{4.15}$$

This results in a $M \times M$ matrix $\mathbf{\Lambda}$ with real eigenvalues λ_k and a $M \times M$ matrix $\mathbf{E}$ with complex eigenvectors E_m^k, where $m = 1 \ldots M$ are the spatial locations and $k = 1 \ldots K$ represent the modes. As usual, only the first K eigenvalues are non zero so that the effective dimension of $\mathbf{\Lambda}$ is $K \times K$ and that of $\mathbf{E}$ is $M \times K$. As in standard EOF analysis, the kth complex EOF mode has a fraction of the total field variance associated with it, which is proportional to the eigenvalue λ_k in the way given by equation (2.13).

The eigenvectors E_m^k are the spatial EOF patterns as in traditional EOF analysis (see equation (2.10)), but in this case they are complex, that is, they consist of a real and an imaginary part. As such, they can be expressed in terms of a spatial amplitude B_m^k and a spatial phase Θ_m^k:

$$E_m^k \;=\; B_m^k \; e^{\,i\,\Theta_m^k} \tag{4.16}$$

Principal components are also constructed as in traditional EOF analysis, by projecting the spatial EOFs onto the original data set, as in equations (2.11)- (2.12). The resulting time series $A^k(t)$ are also complex, that is, they consist of a temporal amplitude $C^k(t)$ and a temporal phase $\Psi^k(t)$:

$$A^k(t) \;=\; C^k(t) \; e^{\,i\,\Psi^k(t)} \tag{4.17}$$

The four mentioned measures (spatial amplitude B_m^k, spatial phase Θ_m^k, temporal amplitude $C^k(t)$ and temporal phase $\Psi^k(t)$) constitute a general description of possible moving features in the original field $F_m(t)$ and can be determined from the eigendecomposition without any assumption on the form of $F_m(t)$. Since these features are defined in the space/time domain, their interpretation does not suffer in the presence of cyclo-stationary signals. In fact, they give a clear description of periodicities, if they are present in the data.
The spatial amplitude B_m^k shows the spatial distribution of variability associated with each eigenmode and may be interpreted as the EOF pattern in standard EOF analysis. It is defined as:

$$B_m^k \;=\; \left(E_m^{k*}\, E_m^k\right)^{1/2} \tag{4.18}$$

The spatial phase Θ_m^k shows the relative phase of the fluctuations among the various spatial locations where F is defined. This measure, for which an arbitrary initial value must be selected, varies continuously between 0° and 360°. As such, it is no longer restricted to the two possible values it can take in standard EOF analysis (0° or 180°). The spatial phase Θ_m^k is defined as:

$$\Theta_m^k \;=\; \arctan\left(\frac{\Im\left\{E_m^k\right\}}{\Re\left\{E_m^k\right\}}\right) \tag{4.19}$$

The temporal amplitude $C^k(t)$ provides a measure of the temporal variability in the magnitude of the structure of the mode and may be interpreted as the PC in standard EOF analysis. It is obtained from:

$$C^k(t) \;=\; \left(A^{k*}(t)\, A^k(t)\right)^{1/2} \tag{4.20}$$

The temporal phase $\Psi^k(t)$ describes the temporal variation of the phase associated with periodicities in field F. It is defined as:

$$\Psi^k(t) \;=\; \arctan\left(\frac{\Im\left\{A^k(t)\right\}}{\Re\left\{A^k(t)\right\}}\right) \tag{4.21}$$

For a given mode k, the evolution of the phase $\Psi^k(t)$ with time reveals information about the existence of quasi-periodicities in the field. In fact, if $\Psi^k(t)$ increases monotonically from 0° to 360° over any 360° interval, it can be inferred that a certain cyclicity exists in the data (see Venegas et al., 1998; Tourre et al., 1999b, for examples of $\Psi(t)$).

As in traditional EOF analysis, a compressed and less noisy version of the complex field $F_m(t)$ can be reconstructed as the sum of the contributions of the leading H empirical orthogonal functions ($H < K$):

$$F_m(t) \;=\; \sum_{k=1}^{H} E_m^{k*}\, A^k(t) \tag{4.22}$$

The reconstructed version of the *real* field of observations $\phi_m(t)$ is thus recovered by taking the real part of equation (4.22).

If the fluctuations in field F are suitably simple, the four measures described here are considerably easy to interpret. However, as the complexity of the field increases beyond several propagating irregular structures, fluctuating at irregular intervals, the interpretations are in general no longer easy. In such a case, the effectiveness of the CEOF decomposition may be enhanced by pre-filtering the data series *before* performing the analysis, that is, by concentrating the attention on a specific frequency band. However, this implies a subjective choice of a timescale of focus. This choice may be made based on the specific objective of the study or on a prior spectral analysis on the data. Examples of pre-filtering of the data in combination with CEOF analysis can be found in Venegas et al. (1998) and Tourre et al. (1999b).

A more intuitive way of presenting the results from a CEOF analysis, instead of the presentation of the four measures described above, involves the reconstruction of the original field based on only one mode of the decomposition (usually the first). Using equation (4.22) with $H = 1$ and taking the real part only, a series of N maps of the reconstructed signal can be computed, one for each time t. We can thus make composites (averages) of these maps at those times t that correspond to specific values of the temporal phase $\Psi(t)$. A reasonable choice would be to average maps at times t for which $\Psi = 0°$, $90°$, $180°$, $270°$ and $360°$ (that is, 5 time steps spanning one complete cycle). By doing this, we obtain a sequence of 5 composites (averaged spatial patterns) which we can consider as instantaneous "pictures" or "snapshots" of the signal of interest at different phases of one cycle (note that, in fact, the composite for $\Psi = 360°$ is redundant since it is identical to that for $\Psi = 0°$). This sequence of consecutive snapshots provides information on the time evolution of the signal associated with the reconstructed mode during one "typical" oscillation. A careful inspection of such a sequence of maps allows for the detection of propagating structures simply by following the evolution of the structures through the consecutive pictures.

As an extension of this method, a Combined CEOF analysis may also be performed with the purpose of investigating the joint variability of two covarying fields. In this case, the definition of the original field $\phi_m(t)$ is changed so as to include the time series of the two normalized fields in the manner described in section 2.2 (matrix (2.28)). Examples of the usage of the CEOF method can be found in Rasmusson et al. (1981); Barnett (1983); Horel (1984); Latif and Barnett (1994); Lanzante (1996); Venegas et al. (1998); White et al. (1998); Mysak and Venegas (1998); Enfield and Mestas-Nuñez (1999); Tourre et al. (1999b); von Storch and Zwiers (1999); and White and Cayan (2000).

4.4 Multichannel Singular Spectrum Analysis (MSSA)

Multichannel Singular Spectrum Analysis (MSSA) is a generalization of the SSA method when both *space* and *lag* vary in the data matrix **F**. As such, MSSA is simply the multivariate version of SSA. Performing MSSA is like performing SSA simultaneously for a number of time series. The name "Multichannel" comes from the introduction of more than one time series in the analysis.

MSSA is mathematically equivalent to the Extended EOF (EEOF) method described in section 4.1. In practice, these two techniques only differ in their domain of application. In general, MSSA deals with more temporal degrees of freedom (lags) than spatial ones, hence it is particularly focused on the time coordinate, allowing the development of interesting spectral properties. In EEOF analysis, however, the data matrix **F** contains only a few lags and a large number of spatial locations. Both methods seek the dominant space-time patterns of the analyzed signals, taking into account both their temporal and spatial correlations. In MSSA, the data matrix contains both space and time information. The eigenvectors obtained from the decomposition of the covariance matrix $\mathbf{R_{FF}}$ are now sequences of spatial patterns and are called Space-Time-EOFs (ST-EOFs) in analogy with the T-EOFs in the univariate SSA. Similarly, the principal components in MSSA are commonly called Space-Time-PCs (ST-PCs)

Like in EEOF analysis, the MSSA approach encounters severe dimensional limitations when analyzing large data sets. The introduction of multiple "channels" (spatial locations) in the estimation of the covariance matrix requires the decomposition of a matrix in the time, spatial and lag domains. For time series of length N measured at M spatial points or channels, and using L lags, the method requires the decomposition of a huge $(L+1)*M \times N-L$ covariance matrix (see section 4.1 for details). To avoid this problem, the multivariate data set may be decomposed into a lower-dimensional data set by means of a conventional EOF analysis, prior to the application of the MSSA. By doing this, however, we are introducing some of the limitations of the classical EOF analysis discussed in section 2.1.5. In this sense, the usefulness of the MSSA is limited to data sets with relatively small spatial extension.

Since the MSSA procedure is identical to that of Extended EOF analysis, a complete and detailed description can be found in section 4.1. For further details on the MSSA technique, see for example Keppenne and Ghil (1993); Plaut and Vautard (1994); Allen and Robertson (1996); Moron et al. (1998); von Storch and Zwiers (1999); and Vautard (1999).

4.5 Multi Taper Method - Singular Value Decomposition (MTM-SVD)

The Multi Taper Method - Singular Value Decomposition (MTM-SVD) approach is a multivariate frequency-domain decomposition technique developed by Mann and Park (1994, 1996, 1999). The MTM-SVD method seeks to identify statistically significant narrow-band oscillations (which may be modulated in phase and amplitude) that are correlated among a large number of time series (locations). It exploits the Multi Taper Method (MTM) of spectral analysis, which we have described in section 3.1, in combination with the singular value decomposition approach described in section 2.1.4.

Standarized time series of field $F_m(t)$ are first computed using equation (2.1), where $t = 1 \dots N$ span the time steps and $m = 1 \dots M$ span the spatial locations. Each time series is first transformed from the time to the spectral domain by using the univariate MTM approach described in section 3.1. Briefly recalling the MTM procedure, for each of the M time series we calculate S independent eigenspectra $Y_m^s(f)$ by multiplying the time series by a family of S Slepian tapers, like in equation (3.2). Appropriate values of p and S are chosen as in the univariate case (see section 3.1 for details). However, instead of averaging the S eigenspectra together for each time series, as we did in section 3.1, the MTM-SVD approach seeks to retain the independent statistical information provided by each of the S eigenspectra by finding an optimal linear combination of them. This optimal linear combination is such that it maximizes the variance (over all locations) explained by the dominant signal in each frequency band of the decomposition.

We first organize the values of the S eigenspectra $Y_m^s(f)$ for the M time series and for each frequency f into a matrix $\mathbf{Y}(f)$ of dimension $M \times S$. Note that matrix $\mathbf{Y}(f)$ is a function of frequency f, which means that we construct one matrix $\mathbf{Y}(f)$ *for each frequency* f of the Fourier decomposition of the eigenspectra $Y_m^s(f)$. That implies the construction of as many matrices $\mathbf{Y}(f)$ as frequencies f are included in the interval $[0, f_n]$, where f_n is the Nyquist frequency (the highest detectable frequency determined by the sampling interval Δ between data points). For example, matrix $\mathbf{Y}(f_o)$ for a given frequency f_o has the following form:

$$\mathbf{Y}(f_o) = \begin{bmatrix} Y_1^1(f_o) & Y_1^2(f_o) & \dots & Y_1^S(f_o) \\ Y_2^1(f_o) & Y_2^2(f_o) & \dots & Y_2^S(f_o) \\ \dots & \dots & \dots & \dots \\ Y_M^1(f_o) & Y_M^2(f_o) & \dots & Y_M^S(f_o) \end{bmatrix} \tag{4.23}$$

We then perform a complex singular value decomposition *on each matrix* $\mathbf{Y}(f_o)$ and obtain several sets of matrices $\mathbf{U}(f_o)$, $\mathbf{\Gamma}(f_o)$ and $\mathbf{V}^\dagger(f_o)$ as in equation (2.19), such that:

$$\mathbf{Y}(f_o) = \mathbf{U}(f_o) * \mathbf{\Gamma}(f_o) * \mathbf{V}^\dagger(f_o) \tag{4.24}$$

Matrix $\mathbf{\Gamma}(f_o)$ is $M \times S$ and contains the singular values $\gamma(f_o)$ in the diagonal. As in section 2.1.4, there are only $K \leq min(M, S)$ non-zero singular values, which defines the maximum number of SVD modes we can determine, so that the effective dimension of matrix $\mathbf{\Gamma}(f_o)$ is $K \times K$.

The columns of the $M \times M$ matrix $\mathbf{U}(f_o)$ are the left singular vectors of $\mathbf{Y}(f_o)$. There are only K useful singular vectors associated with the K non-zero singular values, hence the effective dimension of matrix $\mathbf{U}(f_o)$ is $M \times K$. The K orthogonal M-vectors U_m^k represent the spatial EOF patterns, as in section 2.1.4, except that now their values are complex.

The rows of the $S \times S$ matrix $\mathbf{V}^\dagger(f_o)$ are the right singular vectors of $\mathbf{Y}(f_o)$. The effective dimension of $\mathbf{V}^\dagger(f_o)$ is $K \times S$. The K orthogonal S-vectors V_s^k represent the "spectral EOFs", which we will call "principal modulations" of the frequency-domain decomposition, in analogy with the principal components of the time-domain decomposition. Each principal modulation V_s^k describes the linear combination of projections of the S Slepian tapers that is imposed by the amplitude and phase modulation of the oscillatory signal associated with the kth mode. Using equation (2.22), each matrix $\mathbf{Y}(f_o)$ may be reconstructed as:

$$Y_m^s(f_o) \;=\; \sum_{k=1}^{K} \; U_m^k(f_o) \;\; \gamma_k(f_o) \;\; V_s^k(f_o) \tag{4.25}$$

The main distinction between the MTM-SVD method and the Frequency-domain EOF (FDEOF, see section 4.2) is that the MTM-SVD technique performs a *local* frequency-domain decomposition of S statistically independent eigenspectra, whereas the FDEOF analysis performs a *global* frequency-domain decomposition over all spectral estimates. The MTM-SVD decomposition is performed *locally* in the frequency domain. The term *locally* is interpreted here as meaning "inside a limited frequency interval centered around frequency f_o". It is clear that one singular value decomposition is performed around each resolvable frequency, that is, on each matrix $\mathbf{Y}(f)$ for all $f = 1 \ldots f_n$.

The K singular values $\gamma_k(f)$ scale the amplitude of each mode in this local decomposition and are proportional to the fraction of the variance explained

locally (that is, within a narrow frequency band around frequency f) by the kth mode. The fraction of the local variance explained by the first (the largest) singular value, or "local fractional variance":

$$\text{First Mode Local Fractional Variance (LFV)} \;=\; \frac{\gamma_1^2(f)}{\sum_{k=1}^{K}\ \gamma_k^2(f)} \qquad (4.26)$$

The LFV provides a signal detection parameter that is local in the frequency domain. Using equation 4.26, the LFV accounted for by the first mode can be plotted as a function of frequency, which results in a spectrum-like plot which we call "LFV spectrum". The frequency resolution of the LFV spectrum varies between $\pm f_R$ and $\pm p f_R$, where f_R is the Rayleigh frequency and p is the bandwidth parameter (see section 3.1 for their definitions). That is, the frequency resolution cannot be narrower than the Rayleigh frequency nor greater than the bandwidth corresponding to a uniform average of the S eigenspectra (see Mann and Park, 1999, for further details). Similarly, only variability with periods shorter than $(p f_R)^{-1} = N\Delta/p$ can be confidently distinguished from a secular variation (a trend).

The LFV spectrum provides a powerful parameter for signal detection in the frequency domain, since it shows the fraction of variance explained by the dominant oscillation *in each frequency band*, as a function of frequency. Typically, only the LFV spectrum of the first singular value is used as a signal detection parameter. A peak in the LFV spectrum at a given frequency is indicative of a potentially significant spatiotemporal signal in the dataset, that oscillates at that frequency.

Significance levels in the LFV spectrum are determined through a bootstrapping procedure (Efron, 1990; Mann and Park, 1996). The N maps of the original field F are permuted in time while keeping their spatial structure intact. One thousand permutations of field F are thus generated, which destroys the temporal, but not the spatial, structure of the data field F. The entire MTM-SVD procedure is performed on each of the "randomized" versions of field F and a new LFV spectrum is computed each time. This ensemble of one thousand LFV spectra computed from the re-sampled time series constitutes an estimate of the null distribution of the LFV parameter for spatially correlated coloured noise in the absence of signal. This null distribution is in fact independent of frequency and indistinguishable from that of white noise series with the same underlying spatial correlation. Empirical significance levels are thus obtained by taking the 50%, 90%, 95% and 99% percentiles of this null distribution. For further details on the significance estimation see Mann and Park (1999).

We thus use the LFV spectrum of the first mode with its significance levels

as a parameter for signal detection. By having a look at the LFV spectrum, we determine that, for example, the specific frequency f_o shows statistically significant power (a significant peak in the spectrum). Then, we would like to reconstruct the spatial and temporal patterns of the signal associated with the first mode of the decomposition at frequency f_o. To do this, we use the complex vectors $U_s^1(f_o)$ and $V_s^1(f_o)$, that is, the spatial and spectral EOFs, respectively, obtained from equation (4.25).

The complex vector $U_m^1(f_o)$ of length M is the spatial EOF corresponding to the first mode of the decomposition at f_o. It represents the spatial pattern of the signal at frequency f_o and contains information about the relative phase and amplitude of the signal at all locations m of the multivariate data set. It is the complex equivalent to the first-mode spatial pattern in a traditional EOF decomposition. Rescaling the values of $U_m^1(f_o)$ to account for the standarization performed on the initial data, we can recover the spatial pattern of the first mode signal with the right units as:

$$E_m^1 \;=\; \delta(f_o)\;\; \sigma_m \;\; U_m^1(f_o) \tag{4.27}$$

where σ_m is the standard deviation calculated as in equation (2.3). The factor $\delta(f_o) = 2$ for frequencies $f_o \geq pf_R$, to account for contributions from spectral information at f_o and $-f_o$. For $0 \leq f_o \leq pf_R$ (the secular band), we take $\delta(f_o) = 1$. The complex values in E_m^1 describe the evolution of the EOF spatial pattern over an oscillation of frequency f_o. They may be represented in vectorial form, with the magnitude of the vector indicating relative amplitude and the angle indicating relative phase among the locations m. Examples of the complex spatial pattern E_m^1 can be found in Mann and Park (1994) and Tourre et al. (1999a).

The complex vector $V_s^1(f_o)$ of length S is the spectral EOF corresponding to the first mode of the decomposition at f_o. From this vector we will derive the temporal pattern of the signal at frequency f_o. The temporal pattern of the signal, $A^1(t)$, can be represented as having a dominant oscillation at frequency f_o that suffers amplitude and phase variations on a timescale longer than the oscillation period $1/f_o$, as follows:

$$A^1(t) \;=\; \Re\left\{\alpha(t)\; e^{-i\,2\pi f_o t}\right\} \tag{4.28}$$

where the variable amplitude $\alpha(t)$ represents the slowly varying envelope of the oscillatory signal. The signal in the time domain $A(t)$ and the envelope $\Re\{\alpha(t)\}$ are formally identical for modes referenced to frequency $f_o = 0$, that

is, the secular modes of variability or trends. The slowly varying envelope $\alpha(t)$ at frequency f_o can be obtained by "inverting" the complex vector $V_s^1(f_o)$. This procedure is similar to the complex demodulation, a technique to determine how the signal characteristics at a specific frequency f_o change with time throughout the duration of the time series.

It can be shown that the slowly varying envelope or amplitude $\alpha(t)$ of any oscillatory signal of the form of equation (4.28) centered at frequency f_o can be estimated from a set of eigenspectra $Y_s(f_o)$, $s = 1 \ldots S$ (Park, 1992; Park and Maasch, 1993). In the multivariate case of the MTM-SVD approach, the envelope $\alpha(t)$ of the time domain signal can be reconstructed from the components of the spectral EOF $V_s^1(f_o)$. This reconstruction is not unique and requires additional constraints. The simplest reconstruction is a MTM version of the complex demodulation, that is, a linear combination of the Slepian tapers $a_s(t)$, $s = 1 \ldots S$ and $t = 1 \ldots N$ (Park, 1992; Park and Maasch, 1993):

$$\alpha(t) \; = \; \sum_{s=1}^{S} \; \lambda_s^{-1} \; V_s^1(f_o) \; a_s(t) \tag{4.29}$$

where $V_s^1(f_o)$ is the sth component of the spectral EOF of the first mode and λ_s are the eigenvalues associated with the S Slepian eigentapers (from equation 3.1). Eigenvalues λ_s measure the spectral leakage resistance of the eigentapers a_s. Eigentapers with $\lambda_s \sim 1$ can be used to construct spectral estimates that are resistant to spectral leakage. This reconstruction of $\alpha(t)$ tends to minimize the size of the envelope and as such, it makes $\alpha(t)$ to approach zero at the ends of the time series. Such an inversion is clearly not appropriate for signals in the secular band (trends). An alternative inversion for that case minimizes the first derivative of $\alpha(t)$, that makes the envelope to approach zero slope at the ends of the time series. Such an inversion is more appropriate when there are trends in the time series but it is poorly suited for the description of other features such as rapid changes near the beginning or the end of the series. A third possibility is to minimize the roughness of the envelope using the second derivative of $\alpha(t)$, which constrains neither the mean nor the slope at the ends of the series. A more general data-adaptive means of signal reconstruction in the time domain was suggested by Mann and Park (1996), in which the mean-square multivariate misfit with the raw data is minimized over all possible combinations of the 3 mentioned constraints. This approach removes the subjectivity inherent in the "a priori" choice of the quantity to minimize. For further details about the time-domain signal reconstruction refer to Park (1992); Park and Maasch (1993); and Mann and Park (1999).

We have now obtained the spatial pattern E_m^1 (equation (4.27)) and the temporal pattern $A^1(t)$ (equation (4.28)) of the signal associated with the first mode. They were derived from the spatial EOF (U_m^1) and the spectral EOF (V_s^1) of the decomposition at frequency f_o. We can thus reconstruct the spatiotemporal signal $F_m^1(t)$ (where the super-index 1 indicates that it is the reconstruction of the signal associated with the first mode) for all times and locations as the product of these spatial and temporal patterns as in traditional EOF (equation (2.14)):

$$F_m^1(t) \;=\; E_m^1 \; A^1(t) \tag{4.30}$$

Replacing E_m^1 and $A^1(t)$ by their expressions in equations (4.27) and (4.28), respectively, we get:

$$F_m^1(t) \;=\; \delta(f_o) \;\; \Re\left\{\sigma_m \;\; U_m^1(f_o) \;\; \alpha^1(t) \;\; e^{-i\,2\pi f_o t}\right\} \tag{4.31}$$

This reconstruction of the signal associated with the first mode of the decomposition at frequency f_o can be used to describe the oscillatory pattern in a more intuitive way than the presentation of the spatial and temporal patterns separately. The time-varying amplitude and phase information is more physically provided by presenting a sequence of real-valued patterns (maps) of the signal corresponding to several phases $\Psi(t) = 2\pi f_o t$ of a cycle. Then, for example, a sequence of spatial patterns (maps) of the reconstructed signal can be shown for those times t for which $\Psi(t) = 0°, 90°, 180°, 270°$ and $360°$, which covers a complete oscillation. Maps for $\Psi(t) = 0$ and $\Psi(t) = 360$ are identical. Such a sequence of maps describes the evolution of the signal during an average or "typical" cycle of period $1/f_o$. Further reading and applications of the MTM-SVD method can be found in Mann and Park (1994, 1996); Rajagopalan et al. (1998); Mann and Park (1999); Tourre et al. (1999a); Venegas and Mysak (2000); and Delworth and Mann (2000). Fortran codes to perform an MTM-SVD analysis are provided by Michael Mann at http://www.people.Virginia.edu/~mem6u/mtmsvd.html.

The MTM-SVD technique can also be extended for its application to coupled fields, that is, to more than one field at a time, in order to study the covariability between variables, in the same way as the Combined EOF described in section 2.2. The time series of the two (or more) variables are transformed into the frequency domain using the same S eigentapers, and the resulting eigenspectra of the two fields are concatenated to form matrix $\mathbf{Y}$. The columns of matrix $\mathbf{Y}$ now include the locations of the first field and the

locations of the second field after one another. In this case, it is recommended to weight the eigenspectra of the two fields in matrix $\mathbf{Y}$. The weights are set to be inversely proportional to the number of locations for each field in order to ensure that both data sets are given equal overall weight in the analysis. Similarly to traditional Combined EOF, the spatial eigenvectors now contain a map for each of the two variables, concatenated after one another.

Bibliography

Allen, M. R., and A. W. Robertson, 1996: Distinguishing modulated oscillations from coloured noise in multivariate datasets *Clim. Dyn.* **12** (11), 775–784.

Allen, M. R., and L. A. Smith, 1996: Monte Carlo SSA: detecting irregular oscillations in the presence of coloured noise *J. Climate* **9** (12), 3373–3404.

Barnett, T. P., 1983: Interaction of the Monsoon and Pacific trade wind system at interannual timescales, I, The equatorial zone *Mon. Weather Rev.* **111**, 756–773.

Barnett, T. P., and R. W. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis *Mon. Weather Rev.* **115**, 1825–1850.

Barnston, A. G., and R. E. Livezey, 1987: Classification, seasonality and persistence of low frequency circulation patterns *Mon. Weather Rev.* **115**, 1083–1126.

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

Brillinger, D. R., 1981: *Time series: data analysis and theory. Expanded Edition.* Holden-Day, San Francisco, 540 pp.

Chang, P., L. Ji, and H. Li, 1997: A decadal climate variation in the tropical Atlantic Ocean from thermodynamic air-sea interactions *Nature* **385**, 516–518.

Chelliah, M., and P. Arkin, 1992: Large-scale interannual variability of monthly outgoing longwave radiation anomalies over global tropics *J. Climate* **5**, 371–389.

Cheng, X., G. Nitsche, and J. M. Wallace, 1995: Robustness of low-frequency circulation patterns derived from EOF and Rotated EOF analyses *J. Climate* **8**, 1709–1713.

Cherry, S., 1996: Singular Value Decomposition Analysis and Canonical Correlation Analysis *J. Climate* **9**, 2003–2009.

Cherry, S., 1997: Some comments on singular value decomposition analysis *J. Climate* **10** (7), 1759–1761.

Daubechies, I., 1992: *Ten lectures on Wavelets* Society for Industrial and Applied Mathematics, 357 pp.

Delworth, T. L., and M. E. Mann, 2000: Observed and simulated multidecadal variability in the Northern Hemisphere *Clim. Dyn.* **16**, 661–676.

Deser, C., and M. L. Blackmon, 1993: Surface climate variations over the North Atlantic Ocean during winter: 1900-1989 *J. Climate* **6**, 1743–1753.

Deser, C., and M. S. Timlin, 1997: Atmosphere-ocean interaction on weekly timescales in the North Atlantic and Pacific *J. Climate* **10** (3), 393–408.

Dettinger, M. D., M. Ghil, and C. L. Keppenne, 1995: Interannual and interdecadal variability in United States surface air temperatures, 1910-87 *Clim. Change* **31**, 35–66.

Efron, B., 1990: *The Jackknife, the Bootstrap and other resampling plans* Society for Applied and Industrial Mathematics, Philadelphia, 92 pp.

Emery, W. J., and R. E. Thomson, 1998: *Data analysis methods in physical oceanography* Pergamon, Elsevier, New York, 634 pp.

Enfield, D. B., and A. M. Mestas-Nuñez, 1999: Multiscale variabilities in global sea surface temperatures and their relationships with tropospheric climate patterns *J. Climate* **12** (9), 2719–2733.

Frankignoul, C., F. Bonjean, and G. Reverdin, 1996: Interannual variability of surface currents in the tropical Pacific during 1987-1993 *J. Geophys. Res.* **101** (C2), 3629–3647.

Gamage, N., and W. Blumen, 1993: Comparative analysis of low-level cold fronts: wavelet, Fourier and empirical orthogonal function decompositions *Mon. Weather Rev.* **121**, 2867–2878.

Graham, N. E., J. Michaelsen, and T. P. Barnett, 1987: An investigation of the El Nino-Southern Oscillation cycle with statistical models. I. Predictor field characteristics *J. Geophys. Res.* **92** (C13), 14251–14270.

Gu, D., and S. G. H. Philander, 1997: Secular changes of annual and interannual variability in the tropics during the past century *J. Climate* **8**, 864–876.

Horel, J. D., 1984: Complex Principal Component Analysis: theory and examples *J. Clim. Appl. Meteorol.* **23**, 1660–1673.

Houghton, R., and Y. Tourre, 1992: Characteristics of low frequency sea surface temperature fluctuations in the tropical Atlantic *J. Climate* **5**, 765–771.

Keppenne, C. L., and M. Ghil, 1992: Adaptive filtering and prediction of the Southern Oscillation *J. Geophys. Res.* **97** (D18), 20449–20454.

Keppenne, C. L., and M. Ghil, 1993: Adaptive filtering and prediction of noisy multivariate signals: an application to atmospheric angular momentum *Int. J. Bifurc. and Chaos* **3**, 625–634.

Kousky, V. E., and M. T. Kayano, 1994: Principal modes of outgoing longwave radiation and 250-mb circulation for the South American sector *J. Climate* **7** (7), 1131–1143.

Kuo, C., C. Lindberg, and D. J. Thomson, 1990: Coherence established between atmospheric carbon dioxide and global temperature *Nature* **343**, 709–713.

Lanzante, J. R., 1984: A rotated eigenanalysis of the correlation between 700 mb heights and sea surface temperature in the Pacific and Atlantic *Mon. Weather Rev.* **112**, 2270–2280.

Lanzante, J. R., 1996: Lag relationships involving tropical sea surface temperatures *J. Climate* **9**, 2568–2578.

Latif, M., and T. P. Barnett, 1994: Causes of decadal climate variability over the North Pacific and North America *Science* **266**, 634–637.

Lau, K. M., and P. H. Chan, 1986: Aspects of the 40-50 day oscillation during the northern summer as inferred from outgoing longwave radiation *Mon. Weather Rev.* **114** (7), 1354–1367.

Lau, K. M., and H. Weng, 1995: Climate signal detection using wavelet transform: How to make a time series sing *Bull. Am. Meteorol. Soc.* **76**, 2391–2402.

Liu, P. C., 1994: Wavelet spectrum analysis and ocean wind waves in *Wavelets in Geophysics*, edited by E. Foufoula-Georgiou and P. Kumar pp. 151–166 Academic Press.

Mann, M. E., and J. Park, 1993: Spatial correlations of interdecadal variation in global surface temperatures *Geophys. Res. Lett.* **20**, 1055–1058.

Mann, M. E., and J. Park, 1994: Global-scale modes of surface temperature variability on interannual to century timescales *J. Geophys. Res.* **99**, 25819–25833.

Mann, M. E., and J. Park, 1996: Joint spatiotemporal modes of surface temperature and sea level pressure variability in the northern hemisphere during the last century *J. Climate* **9**, 2137–2162.

Mann, M. E., and J. Park, 1999: Oscillatory spatiotemporal signal detection in climate studies: A multiple-taper spectral domain approach *Advances in Geophys.* **41**, 1–131.

Mestas-Nuñez, A. M., and D. B. Enfield, 1999: Rotated global modes of non-ENSO sea surface temperature variability *J. Climate* **12** (9), 2734–2746.

Meyers, S. D., B. G. Kelly, and J. J. O'Brien, 1993: An introduction to wavelet analysis in oceanography and meteorology with application to the dispersion of Yanai waves *Mon. Weather Rev.* **121**, 2858–2866.

Moron, V., R. Vautard, and M. Ghil, 1998: Trends, interdecadal and interannual oscillations in global sea surface temperature *Clim. Dyn.* **14** (7-8), 545–569.

Mysak, L. A., and S. A. Venegas, 1998: Decadal climate oscillations in the Arctic: A new feedback loop for atmosphere-ice-ocean interactions *Geophys. Res. Lett.* **25** (19), 3607–3610.

Park, J., 1992: Envelope estimation for quasi-periodic signals in noise: a multitaper approach in *Statistics in the Environmental and Earth Sciences*, edited by A. T. Walden and P. Guttorp pp. 189–219 Edward Arnold, London.

Park, J., and K. A. Maasch, 1993: Plio-Pleistocene time evolution of the 100-kyr cycle in marine paleoclimate records *J. Geophys. Res.* **98**, 447–461.

Park, J., C. R. Lindberg, and F. L. Vernon III, 1987: Multitaper spectral analysis of high-frequency seismograms *J. Geophys. Res.* **92**, 12675–12684.

Peixoto, J. P., and A. H. Oort, 1992: *Physics of Climate* American Institute of Physics, New York, 520 pp.

Peng, S., and J. Fyfe, 1996: The coupled patterns between sea level pressure and sea surface temperature in the mid-latitude North Atlantic *J. Climate* **9** (8), 1824–1839.

Percival, D. B., and A. T. Walden, 1993: *Spectral analysis for physical applications. Multitaper and conventional univariate techniques* Cambridge University Press, Cambridge, United Kingdom, 583 pp.

Peterson, R. G., and W. B. White, 1998: Slow oceanic teleconnections linking the Antarctic Circumpolar Wave with the tropical El Niño-Southern Oscillation *J. Geophys. Res.* **103** (C11), 24573–24583.

Plaut, G., and R. Vautard, 1994: Spells of low-frequency oscillations and weather regimes in the Northern Hemisphere *J. Atmos. Sci.* **51** (2), 210–236.

Preisendorfer, R. W., 1988: *Principal Component Analyses in Meteorology and Oceanography* Elsevier, Amsterdam, 425 pp.

Rajagopalan, B., M. E. Mann, and U. Lall, 1998: A multivariate frequency-domain approach to long-lead climatic forecasting *Weather and Forecasting* **13**, 58–74.

Rasmusson, E. M., P. A. Arkin, W. Y. Chen, and J. B. Jalickee, 1981: Biennial variations in surface temperature over the United States as revealed by singular decomposition *Mon. Weather Rev.* **109**, 181–192.

Richman, M. B., 1986: Rotation of principal components *Int. J. Climatol.* **6**, 293–335.

Shen, Z., W. Wang, and L. Mei, 1994: Finestructure of wind waves analyzed with wavelet transform *J. Phys. Oceanogr.* **24**, 1085–1094.

Slepian, D., 1978: Prolate spheroidal wave function, Fourier analysis and uncertainty, V, The discrete case *Bell Syst. Tech. J.* **57**, 1371–1429.

Tanimoto, Y., N. Iwasaka, K. Hanawa, and Y. Toba, 1993: Characteristic variations of sea surface temperature with multiple time scales in the North Pacific *J. Climate* **6**, 1153–1160.

Thomson, D. J., 1982: Spectrum estimation and harmonic analysis *IEEE Proc.* **70**, 1055–1096.

Thomson, D. J., 1995: The seasons, global temperature and precession *Science* **268**, 59–68.

Torrence, C., and G. P. Compo, 1998: A practical guide to Wavelet Analysis *Bull. Am. Meteorol. Soc.* **79** (1), 61–78.

Tourre, Y. M., and W. B. White, 1995: ENSO signals in global upper-ocean temperature *J. Phys. Oceanogr.* **25** (6), 1318–1331.

Tourre, Y. M., and W. B. White, 1997: Evolution of the ENSO signal over the Indo-Pacific domain *J. Phys. Oceanogr.* **27** (5), 683–96.

Tourre, Y. M., B. Rajagopalan, and Y. Kushnir, 1999a: Dominant patterns of climate variability in the Atlantic Ocean during the last 136 years *J. Climate* **12**, 2285–2299.

Tourre, Y. M., Y. Kushnir, and W. B. White, 1999b: Evolution of interdecadal variability in sea level pressure, sea surface temperature and upper ocean temperature over the Pacific Ocean *J. Phys. Oceanogr.* **29** (7), 1528–1541.

Trenberth, K. E., and W.-T. K. Shin, 1984: Quasibiennial fluctuations in sea level pressure over the Northern Hemisphere *Mon. Weather Rev.* **112**, 761–777.

Vautard, R., 1999: Patterns in time: SSA and MSSA in *Analyses of Climate Variability - Applications of Statistical Techinques. Second Edition*, edited by H. von Storch and A. Navarra pp. 265–286 Springer-Verlag, Berlin.

Vautard, R., P. Yiou, and M. Ghil, 1992: Singular-spectrum analysis: a toolkit for short, noisy chaotic signals *Physica D* **58**, 95–126.

Venegas, S. A., and L. A. Mysak, 2000: Is there a dominant timescale of natural climate variability in the Arctic? *J. Climate* **13** (19), 3412–3434.

Venegas, S. A., L. A. Mysak, and D. N. Straub, 1996: Evidence for interannual and interdecadal climate variability in the South Atlantic *Geophys. Res. Lett.* **23** (19), 2673–2676.

Venegas, S. A., L. A. Mysak, and D. N. Straub, 1997: Atmosphere-ocean coupled variability in the South Atlantic *J. Climate* **10** (11), 2904–2920.

Venegas, S. A., L. A. Mysak, and D. N. Straub, 1998: An interdecadal climate cycle in the South Atlantic and its links to other ocean basins *J. Geophys. Res.* **103** (C11), 24723–24736.

von Storch, H., 1999: Spatial patterns: EOFs and CCA in *Analyses of Climate Variability - Applications of Statistical Techinques. Second Edition*, edited by H. von Storch and A. Navarra pp. 231–263 Springer-Verlag, Berlin.

von Storch, H., and C. Frankignoul, 1998: Empirical modal decomposition in coastal oceanography in *The Sea. Volume 10. The global coastal oceans. Processes and Methods*, edited by K. Brink and A. Robinson pp. 419–455 John Wiley and Sons, New York.

von Storch, H., and A. Navarra, eds., 1999: *Analyses of Climate Variability - Applications of Statistical Techinques. Second Edition* Springer-Verlag Berlin.

von Storch, H., and F. Zwiers, 1999: *Statistical analysis in climate research* Cambridge University Press, Cambridge, United Kingdom, 484 pp.

Wallace, J. M., and R. E. Dickinson, 1972: Empirical orthogonal representation of time series in the frequency domain. Part I: theoretical considerations *J. Appl. Meteor.* **11** (6), 887–892.

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

Wang, B., and Y. Wang, 1996: Temporal structure of the Southern Oscillation as revealed by waveform and wavelet analysis *J. Climate* **9**, 1586–1598.

Weare, B. C., and J. S. Nasstrom, 1982: Examples of Extended Empirical Orthogonal Function Analyses *Mon. Weather Rev.* **110**, 481–485.

Weng, H., and K.-M. Lau, 1994: Wavelets, period doubling, and time-frequency localization with application to organization of convection over the tropical western Pacific *J. Atmos. Sci.* **51** (17), 2523–2541.

Werner, P. C., and H. von Storch, 1993: Interannual variability of central european mean temperature in January-February and its relation to large-scale circulation *Clim. Res.* **3**, 195–207.

White, W. B., and D. R. Cayan, 2000: A global el nino-southern oscillation wave in surface temperature and pressure and its interdecadal modulation from 1900 to 1997 *J. Geophys. Res.-Oceans* **105** (C5), 11223–11242.

White, W. B., and R. G. Peterson, 1996: An Antarctic circumpolar wave in surface pressure, wind, temperature and sea-ice extent *Nature* **380**, 699–702.

White, W. B., C. Yi, and T. Chang-Kou, 1998: Coupling of biennial oceanic Rossby waves with the overlying atmosphere in the Pacific Basin *J. Phys. Oceanogr.* **28** (6), 1236–1251.

Yi, D., L. A. Mysak, and S. A. Venegas, 1999: Decadal to interdecadal fluctuations of Arctic sea ice cover and the atmospheric circulation during 1954-1994 *Atmos.-Ocean* **37** (4), 389–415.

Zorita, E., V. Kharin, and H. von Storch, 1992: The atmospheric circulation and sea surface temperature in the North Atlantic area in winter: their interaction and relevance for Iberian precipitation *J. Climate* **5** (10), 1097–1108.