A Primer for EOF Analysis of Climate Data

A. Hannachi

Department of Meteorology, University of Reading
Reading RG6 6BB, U.K.
email: A.Hannachi@reading.ac.uk

March 24, 2004

Contents

1 Introduction

Climate is regarded as the aggregation of (random) daily weather, and as pointed out by Lorenz (1970), climate is what we expect but weather is what we get. As such, climate is considered as the long term statistics of weather. Climate variations are also the result of exceedingly complex nonlinear interactions between very many degrees of freedom or modes. Despite being the statistics, or averages, of weather climate is still characterised by nonlinearity and high dimensionality. Consequently, a challenging task is to find ways to redue the dimensionality of the system and find the most important patterns explaining the variations. In the last several decades, meteorologists have in fact put major efforts in extracting important patterns from measurements of atmospheric variables. As a result Empirical Orthogonal Functions (EOF) technique has become the most widely used way to do this and the present primer is a guide for the same purpose.

The original purpose of EOFs was to reduce the large number of variables of the original data to a few variables, but without compromising much of the explained variance. Lately, however, EOF analysis has been used to extract individual modes of variability such as the Arctic Oscillation (AO), (Thompson and Wallace 1998, 2000) known as teleconnections (Angström 1935; Bjerknes 1969; Wallace and Gutzler 1981, etc.). EOFs are presented in section 2.

The simplicity and the analytic derivation of EOFs are the main reasons behind its popularity in atmospheric science. The physical interpretability of the obtained patterns is, however, a matter of controversy because of the strong constraints satisfied by EOFs, namely orthogonality in both space and time. Physical modes such as normal modes (Simmons et al. 1983) are not in general orthogonal. This shortcoming has led to the development of rotated empirical orthogonal functions (REOFs), see Richman (1986). REOFs yields in general localised structures by compromising some of the EOFs properties such as orthogonality. REOFs will be presented in section 3.

EOFs and REOFs are mainly based on using the spatial correlation of the field, an important feature of climate data. Auto- and cross-correlation in time between grid points are, however, ignored in those techniques. Extended EOFs (Weare and Nasstrom, 1982) is a technique that attempts to incorporate both the spatial and the temporal correlation. The method has since become a useful tool to extract dynamical structure, trends and oscillations, and to filter data. EEOFs will be presented in section 5.

2 Structure of climate data

2.1 Data matrix

Gridded climate data normally come as an array containing for each vertical level a three-dimensional, two-dimensional in space and one-dimensional in time, field F. The latter is a function of time t, latitude θ, and longitude ϕ. We suppose that the horizontal coordinates are discretised to yield latitudes θ_j, $j = 1, \ldots, p_1$, and longitudes ϕ_k, $k = 1, \ldots p_2$, and similarly for time, i.e. t_i, $i = 1, \ldots n$. This yields a total number of grid points $p = p_1 p_2$. The discretised field reads:

$$F_{ijk} = F\left(t_i, \theta_j, \phi_k\right) \tag{1}$$

with $1 \leq i \leq n$, $1 \leq j \leq p_1$, and $1 \leq k \leq p_2$. It is in general tedious and memory-consuming to process three- and higher-dimensional arrays such as F. We therefore transform F into a two-dimensional array: the *data matrix* X where the two spatial dimensions are concatenated together. This is achieved using the MATLAB cammand:

```
>> [n, p1, p2] = size (F);
>> X = reshape (F, n, p1*p2);
```

We now suppose that we have a gridded data set composed of a space-time field $X(t, s)$ representing the value of the field X, such as SLP, at time t and spatial position s. The value of the field at discrete time t_i and grid point s_j is noted x_{ij} for $i = 1, \ldots, n$ and $j = 1, \ldots p = p_1 p_2$. The observed field is then represented by the data matrix:

$$X = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{np} \end{pmatrix} \tag{2}$$

If we denote by $\overline{x}_{.j}$ the time average of the field[1] at the j'th grid point, i.e.

$$\overline{x}_{.j} = \frac{1}{n} \sum_{k=1}^{n} x_{kj} \tag{3}$$

then the climatology of the field is defined by

$$\overline{\mathbf{x}} = (\overline{x}_{.1}, \ldots, \overline{x}_{.p})$$

[1] The seasonal and other external, e.g. diurnal cycles are supposed to have been removed from the data.

The anomaly field, or departure from the climatology is defined at (t, s) by:

$$x'_{ts} = x_{ts} - \overline{x}_{.s}$$

or in matrix form:

$$X' = X - \mathbf{1}\overline{\mathbf{x}} = \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) X \tag{4}$$

where $\mathbf{1} = (1, \ldots 1)^T$ is the (column) vector containing n ones, and I is the $n \times n$ identity matrix. If X is the original data matrix, then the MATLAB commands to calculate the anomaly are:

```
>> [n,p] = size (X);
>> Xbar = mean (X,1);
>> Xprime = X - ones(n,1) * Xbar;
```

2.2 Area weighting

Most climate data whether observed or model-simulated are in general non-uniformly distributed over the Earth surface. For example, if the data are provided on a grid with 5^olat$\times 5^o$lon resolution, then clearly the distribution of the data will be denser poleward. This non-uniform distribution can influence the structure of the computed EOFs. In order to avoid the effect of this geometrical artifact we normally weigh the data prior to analysing them. The simplest and most useful way is to weigh each data point by the local area of its location. Hence each datum is weighted by the cosine of its latitude. Let us designate by θ_k the latitude of the k'th grid point, $k = 1, \ldots p$, and D_θ the diagonal matrix:

$$D_\theta = \text{Diag } [\cos\theta_1, \ldots, \cos\theta_p] \tag{5}$$

Then the weighted anomaly matrix is

$$X_w = X' D_\theta. \tag{6}$$

3 Empirical orthogonal functions

3.1 Historical overview

Empirical Orthogonal Function technique aims at finding a new set of variables that capture most of the observed variance from the data through a linear combination of the original variables. Empirical Orthogonal Functions (EOFs) have been introduced in atmospheric science since the early 50's, e.g. Obukhov (1947, 1960), Fukuoka

(1951), Lorenz (1956), see also Craddock (1973) for a discussion on relevant problems in analysing multivariate data in meteorology. The EOF terminology is due to Lorenz (1956) who applied it in a forecasting project at the Massachusetts Institute of Technology. Since then EOFs have become popular analysis tools in climate research.

EOF techniques are deeply rooted in statistics, and go back to Hotelling (1933) who introduced principal component analysis (PCA), another name for EOFs. A review of PCA/EOFs can be found in Kutzbach (1967). EOFs, however, are not restricted to multivariate statistics or atmospheric sciences. They extend to the analysis of stochastic fields in the mathematical literature where they are known under the name Karhunen-Loève basis functions (Loève 1978). Detailed analyses of EOFs can be found for example in Wilks (1995), von Storch and Zwiers (1999), and Jolliffe (2002).

3.2 Derivation and use of EOFs

3.2.1 Derivation of EOFs

We present below a brief description of EOFs, and for more details the reader is referred, for example, to Kutzbach (1967), Wilks (1995), Storch (1993), Storch and Zwiers (1999), and Joliffe (2002). Once the anomaly data matrix (4) or its weighted version (6) is determined, the covariance matrix is then defined by:

$$\Sigma = \frac{1}{n-1} X'^T X', \tag{7}$$

which contains the covariance between any pair of grid points. The aim of EOF/PCA is to find the linear combination of all the variables, i.e. grid points, that explains maximum variance. That is to find a direction $\mathbf{a} = (a_1, \ldots, a_p)^T$ such that $X'a$ has maximum variability. Now the variance of the (centered) time series $X'\mathbf{a}$ is

$$\text{var } (X'\mathbf{a}) = \frac{1}{n-1}\|X'\mathbf{a}\|^2 = \frac{1}{n-1}\left(X'\mathbf{a}\right)^T\left(X'\mathbf{a}\right) = \mathbf{a}^T\Sigma\mathbf{a}$$

To make the problem bounded we normally require the vector $\mathbf{a}$ to be unitary. Hence the problem readily yields:

$$\max_{\mathbf{a}} \left(\mathbf{a}^T\Sigma\mathbf{a}\right), \quad \text{s.t. } \mathbf{a}^T\mathbf{a} = 1 \tag{8}$$

The solution to (8) is a simple eigenvalue problem (EVP):

$$\Sigma\mathbf{a} = \lambda\mathbf{a}. \tag{9}$$

By definition the covariance matrix Σ is symmetrical and therefore diagonalisable. The k'th EOF is simply the k'th eigenvector $\mathbf{a}_k$ of Σ after the eigenvalues, and the

corresponding eigenvectors, have been sorted in decreasing order. The covariance matrix is also semidefinite, hence all its eigenvalues are positive. The eigenvalue λ_k corresponding to the k'th EOF gives a measure of the explained variance by $\mathbf{a}_k$, $k = 1, \dots p$. It is usual to write the explained variance in percentage as:

$$\frac{100\lambda_k}{\sum_{k=1}^{p} \lambda_k}\%$$

The projection of the anomaly field X' onto the k'th EOF $\mathbf{a}_k$, i.e. $\mathbf{c}_k = X'\mathbf{a}_k$ is the k'th principal component (PC)

$$c_k(t) = \sum_{s=1}^{p} x'(t,s)a_k(s) \tag{10}$$

3.2.2 Use of EOFs

Because Σ is diagonalisable the set of its eigenvectors forms an orthogonal basis of the p-dimensional Euclidean space, defined with the natural scalar product. So by construction, the Empirical Orthogonal Functions are orthogonal and the PCs uncorrelated. This completely characterises conventional EOFs. The orthogonality property provides a complete basis where the time-varying field can be separated as:

$$X'(t,s) = \sum_{k=1}^{M} c_k(t)a_k(s) \tag{11}$$

This property constitutes, however, a strong constraint that puts limits to the physical interpretability of individual EOFs since in general physical patterns tend to be non-orthogonal (Simmons et al. 1983).

Furthermore, The eigenvalues of Σ are not necessarily distinct and some may have multiplicity greater than one, i.e. degenerate. Therefore, the separation between eigenvalues can be problematic when two or more eigenvalues are degenerate. The investigation of the degeneracy of the covariance matrix spectrum requires a measure of uncertainty of each eigenvalue that reflects sampling and this is quite difficult to get. In practice there are mainly two ways to compute the uncertainty of the eigenvalues and/or the eigenvectors of Σ. The first one, which is used here, is based on a rule of thumb (North et al. 1982):

$$\begin{aligned} \Delta\lambda_k &\approx \lambda_k\sqrt{\tfrac{2}{n}} \\ \Delta\mathbf{a}_k &\approx \tfrac{\Delta\lambda_k}{\lambda_j - \lambda_k}\mathbf{a}_j \end{aligned} \tag{12}$$

where λ_j is the closest eigenvalue to λ_k, and n is the sample size[2]. Alternatively, one can use Monte Carlo simulations. This can be achieved by forming surrogate data by

[2]To be more precise n is the number of independent data in the sample

resampling a part of the data using randomisation[3].

Another drawback of EOFs comes when we use them to reduce the dimensionality of the data, namely the truncation order that can be applied to the sum (11). In practice the truncation order is obtained by fixing the amount of explained variance, e.g. 80%, and choose the leading EOFs that explain altogether this amount of variance.

3.3 Computational aspects

In practice we do not need to solve the EVP (9). We use a powerful tool from linear algebra namely singular value decomposition (SVD), (e.g. Linz and Wang 2003) to factor any $p \times n$ matrix Y (7) as[4]:

$$Y = L\Lambda R^T \tag{13}$$

where L is a $p \times p$ matrix containing the left singular vectors, R is $n \times n$ matrix containing the right singular vectors, and Λ is a diagonal matrix of the same dimension as Y, i.e. $p \times n$, with nonnegative diagonal elements. These elements are the singular values. Note that the maximum number of nonzero singular values is $M = \min(n, p)$, the rank of the matrix Y. The matrices L and R are both orthonormal, or unitary, in the sense that $L^T L = LL^T = I_M$ and similarly for R.

In meteorological application we take $Y = X'^T$ when $p \leq n$. The EOFs are provided by the columns of L and the PCs by the columns of R. The MATLAB application is obtained as

```
>> Fprime = reshape (Xprime, n, p1, p2);
>> [eofs, pcs, var] = eofsvd (Fprime, 1, lat, 20);
```

where an area weighting is applied by the argument '1' and we retained the first 20 EOFs. Note that the first and the second columns of "var" provide respectively the eigenvalues of the covariance matrix and the explained varainces (%) of the EOFs. To plot the spectrum of the data covariance matrix along with their uncertainties use:

```
>> eigenplot (Xprime, var(:,2))
```

[3]An example would be to randomly select a subsample and apply EOFs, then select another subsample etc. This operation can be repeated many times, which yields various realisations of the eigenelements of the resulting covariance matrix from which one can estimate the uncertainties. Another example would be to fix a subset of variables then scramble them by permuting the data, i.e. breaking the chronological order then apply EOFs, and so on.

[4]In (11) we suppose that $p \leq n$, otherwise we get $p - n$ zero singular values. In this case, i.e. when $p > n$ we normally transpose the matrix Y and apply SVD to get efficient (fast) decomposition.

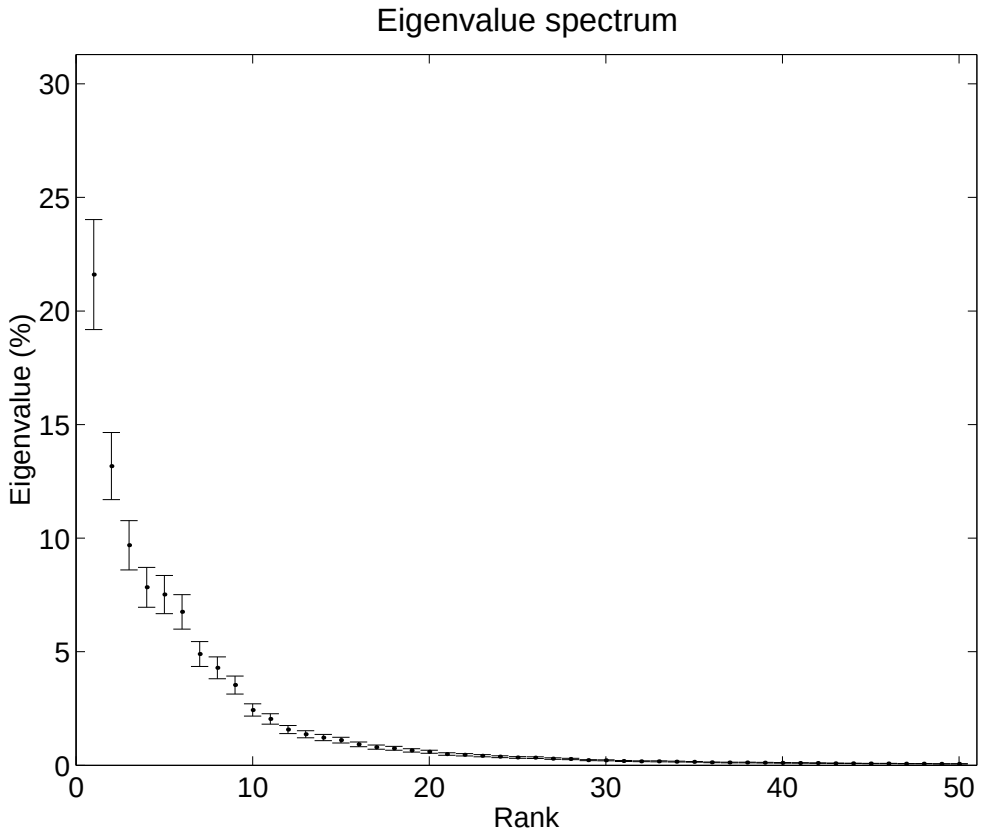

Figure 1. The spectrum of the covariance matrix showing the first 50 eigenvalues (%), i.e. explained variance of the the first 50 EOFs.

3.4 Interpretation of EOFs

By construction EOFs constitute directions of variability with no particular amplitude. Therefore if $\mathbf{l}$ is an EOF, so is $\alpha\mathbf{l}$ for any nonzero α. For convenience, however, they are chosen to be unitary. Also by construction EOFs are stationary structures, i.e. they do not evolve in time. The principal component attached to the corresponding EOF provides the sign and the overall amplitude of the EOF as a function of time. This provides a simplified representation of the state of the field at that time along that EOF. In other words EOFs do not change structure in time, they only change sign and overall amplitude to represent the state of the atmosphere. When EOFs are nondegenerate they can be studied individually. When they are degenerate, however, the separation between them becomes problematic despite being orthogonal and their PCs uncorrelated.

Although EOFs represent patterns that (successively) explain most of the observed variability, their interpretation is, however, not always simple. Physical interpretability especially can be controversial, see e.g. Dommenget and Latif (2002). For physical modes are not necessarily orthogonal (Simmons et al. 1983). The constraints imposed upon EOFs are purely geometric and hence can be non-physical. Furthermore, the EOF structure tends to be domain dependent (Richman 1986) and this adds to the difficulty of their physical interpretability. These arguments constitute the main reasons for attempts to find ways round EOFs. The method of rotated EOFs, which is presented below, constitutes one way.

3.5 Application

We have applied EOFs to the winter monthly SLP over the Northern Hemisphere (NH). The data come from the National Center for Atmospheric Research and National Center for Environmental Prediction (NCAR/NCEP) from 1948 to 2000 on a regular horizontal grid with $2.5^o \times 2.5^o$ resolution. The winter season is defined by December to February (DJF). The seasonal cycle is first calculated as monthly averaged values then removed from the data. Finally a weighting by the cosine of the corresponding latitude is applied. The data over the NH north of 20^oN is used to compute EOFs.

Figure 1 shows the spectrum of the covariance matrix along with their uncertainties using approximation (12). The first eigenvalue seems separated from the rest, but overall the spectrum looks in general continuous, which makes truncation difficult. The first two EOFs (Fig. 2) explain respectively 21% and 13% of the total variance. EOF 1 shows a high over the North Pole and two low centres over the

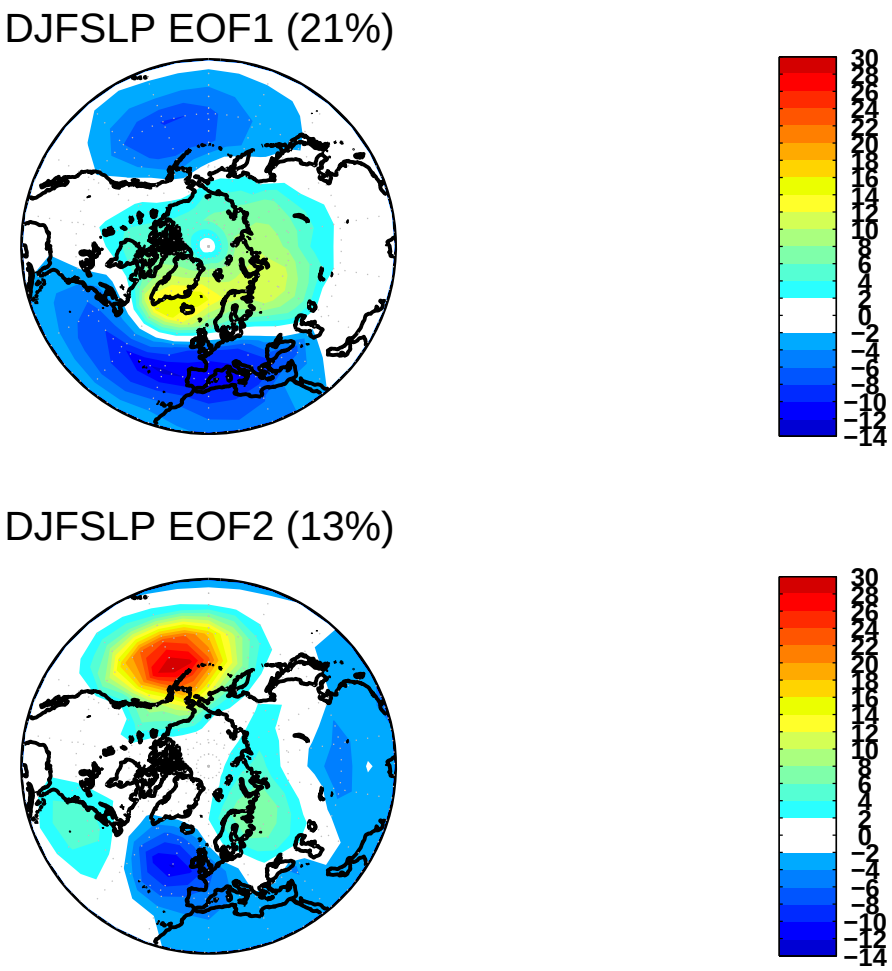

Figure 2. The first two EOFs of the monthly mean sea level pressure.

Mediterranean-North East Atlantic and over the North Pacific. This is the familiar AO mode (Thompson and Wallace 1998; 2000). The second pattern EOF2 shows two separated centres over the North Pacific and North East Atlantic respectively.

Figure 3 shows the first two PCs. The PCs are standardised to have zero mean and unit variance. Times where the PC's cross the horizontal lines ± 1 are arbitrarily chosen to represent particularly high amplitude (or extreme) events. The time series of PC1 (Fig. 3) also shows that during most of the first part of the record the AO was positive as in Fig. 2. In the last part of the record, however, the pattern has mostly changed sign. If now we examine the second component PC2, we can see roughly that the variability in the first part of the record, i.e. until just before the 70's, is lesser than in the remaining part of the record.

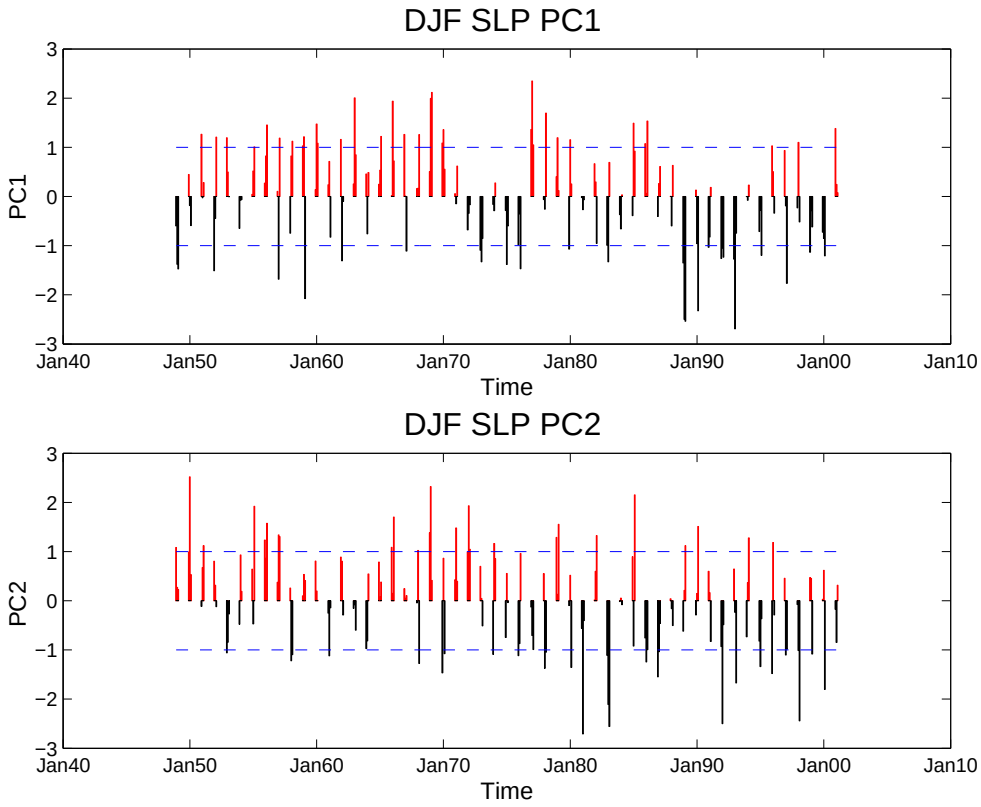

Figure 3. The first two PCs of the monthly mean sea level pressure.

4 Rotated empirical orthogonal functions

4.1 What is it and why?

Rotated EOF (REOF) is a technique simply based on rotating EOFs. REOF techniques have been adopted by atmospheric scientists since the mid-eighties as an attempt to overcome some of the previous shortcomings such as the difficulty of physical interpretability. The technique, however, is much older and goes back to the early fourties (Thurstone 1940, 1947; Caroll 1953). The technique is also known in factor analysis as factor rotation and aims at getting simple structures. In meteorology the objective was to

• alleviate the strong constraints of EOFs, namely orthogonality/uncorrelation of EOFs/PCs, domain dependence of EOF patterns (see e.g., Dommengent and Latif 2002),

• obtain simple structures, and

• be able to physically interpret the patterns.

4.2 Derivation of REOFs

Here one seeks a rotation matrix Q to construct the rotated EOFs U according to:

$$U = LQ \tag{14}$$

where $L = [\mathbf{l}_1, \mathbf{l}_2, \ldots \mathbf{l}_m]$ is the matrix of the leading m EOFs (or loadings). The criterion for choosing the $m \times m$ rotation matrix Q is what constitutes the rotation algorithm, which is expressed by the minimization problem:

$$\min f(LQ) \tag{15}$$

over a specified subset of square rotation matrices Q. The functional f represents the rotation criterion. Various rotation criteria exist in the literature. Richman (1986), for example, lists five simplicity criteria. Broadly speaking there are two large families of rotation namely, orthogonal and oblique rotations. In the former the rotation matrix Q in (14) is chosen to be orthogonal, and the problem is to solve (15) subject to the orthogonality condition:

$$QQ^T = Q^TQ = I. \tag{16}$$

By contrast, in the latter case the rotation matrix Q is non-orthogonal.

The most well known and used rotation algorithm is the varimax criterion (Kaiser 1958). It is an orthogonal rotation based on the criterion:

$$\max \left(\sum_{k=1}^{m} \left[p \sum_{j=1}^{p} u_{ij}^4 - \left(\sum_{j=1}^{p} u_{jk}^2 \right)^2 \right] \right) \tag{17}$$

where $U = (u_{ij})$, and m is the number of EOFs chosen for rotation. The quantity inside the square brackets in (17) is proportional to (spatial) variance of the square of the rotated vector $\mathbf{u}_k = (u_{1k}, \ldots, u_{pk})^T$. Therefore the varimax attempts to simplify the structure of the patterns by tending the loadings towards zero, or ± 1. The MATLAB command to do REOFs is

>> U = roteofs (eofs, m)

4.3 Application

In this manuscript we have applied the varimax rotation to the SLP for various values of the number m of retained EOFs. Fig. 4 shows the first two rotated EOFs for $m = 10$ (first row) and $m = 20$ (2nd row) retained EOFs respectively. For small m ($m \leq 15$) REOF1 shows a clear NAO pattern, a dipolar structure, with the strongest centre located over Greenland and REOF2 shows the Pacific pattern centered around

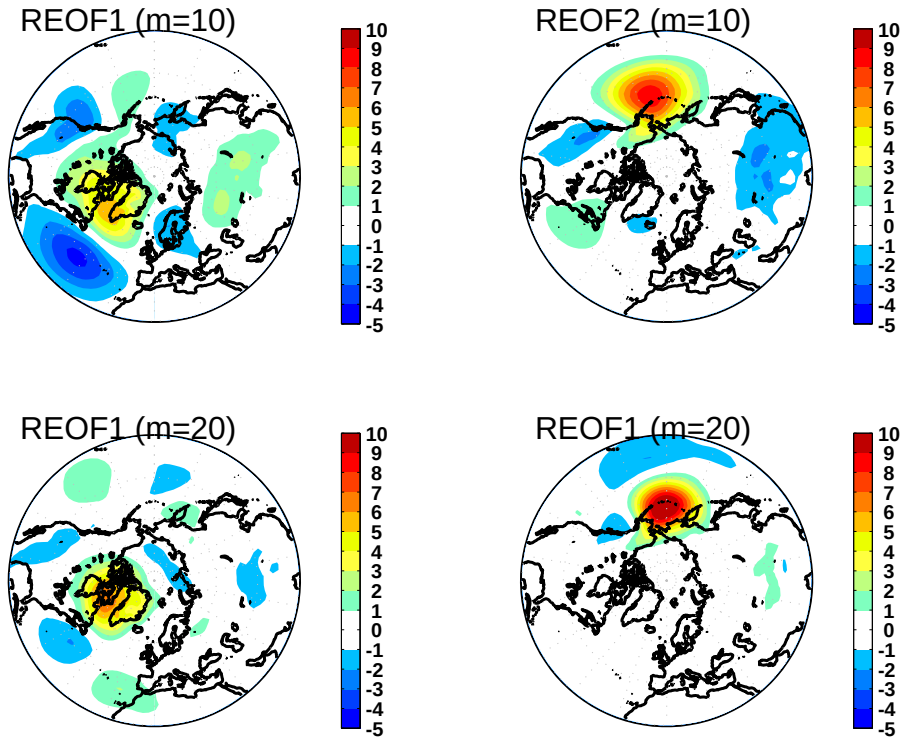

Figure 4. The first two rotated EOFs using 10 (first row) and 20 (2nd row) EOFs respectively.

the Aleutian (Fig 4). As m increases the centre over Western Europe/Eastern Atlantic of REOF1 diminishes and eventually vanishes for $m \geq 25$ and we are left with a very localised centre east of Greenland. Similarly, the Pacific centre of REOF2 becomes very localised. Note that if the EOFS are not area-weighted then REOF1 becomes localised over the North Pole instead of Greenland (not shown). Rotated EOFs therefore yield localised or simple structures, but they are dependent on the number of retained EOFs.

REOFs, like EOFs, also suffer some drawbacks, namely criterion choice, and the number of retained EOFs. In addition the (EOF) property of successive variance maximisation is lost during the rotation process. This yields patterns that do not constitute the main source of variation. Joliffe et al. (2002), for example, pointed to the difficulty in building a criteria to take account of various types of simplicity, e.g. concentration versus equal loadings.

5 Extended empirical orthogonal functions

5.1 Introduction

Weather and climate data have beside their large dimensionality a high correlation in space and also significant correlation in time. EOFs constitute a basic and useful tool to learn about large scale patterns that explain most of the variability. Since EOFs find the combination of variables which explain most of the variability it is implied, in particular, that EOFs make use of the usually observed high correlation in space. The significant correlation in time is, however, not taken into account. Auto- and cross-correlation in time can be very useful for prediction purposes and also for building probabilistic time series models. Extension of EOFs to deal with correlation in time also exist and is presented below.

5.2 Extended EOFs

5.2.1 Historical overview

Extended empirical orthogonal functions (EEOFs) constitute an extension of the traditional EOF technique to deal not only with spatial- but also with temporal correlations observed in weather/climate data. The method has been first introduced by Weare and Nasstrom (1982) who applied it to the 300-mb relative vorticity.

Similar approach has been developed later to deal with dynamical reconstruction of low order chaotic systems by Broomhead and King (1986a,b) who called it *singular system analysis* (SSA) in the unidimensional case. At the same time Fraedrich (1986) also used the same approach to compute dimensions of chaotic attractors from climate data. SSA was also used to find oscillations from climate records (Vautard et al. 1992). It has been extended to deal with multivariate, or multichannel, (MSSA) time series (Broomhead and King 1986a,b) in a way similar to EEOF. MSSA (EEOF) was applied later by Kimoto et al. (1991) and Plaut and Vautard (1994) to find propagating structures from observed 500-mb heights.

5.2.2 Definition of EEOFs

In EEOF the atmospheric state vector at time t, i.e. $\mathbf{x}_t = (x_{t1}, \ldots x_{tp})$, $t = 1, \ldots, n$, used in traditional EOF, is extended to include temporal information as

$$\boldsymbol{x}_t = (x_{t1}, \ldots x_{t+M-1,1}, x_{t2}, \ldots x_{t+M-1,2}, \ldots x_{t,p}, \ldots x_{t+M-1,p}) \tag{18}$$

with $t = 1, \ldots, n - M + 1$. The parameter M used in (18) is known as *window length* or *delay parameter*[5]. The new data matrix takes now the form

$$\mathcal{X} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_{n-M+1} \end{bmatrix} \tag{19}$$

It is now clear from (18) that time is incorporated in the state vector side by side with the spatial dimension. If we denote by

$$\boldsymbol{x}_t^s = (x_{ts}, x_{t+1,s} \ldots x_{t+M-1,s}) \tag{20}$$

then the extended state vector (18) is written in a similar form to the conventional state vector, i.e.

$$\boldsymbol{x}_t = \left(\boldsymbol{x}_t^1, \boldsymbol{x}_t^2, \ldots, \boldsymbol{x}_t^p\right) \tag{21}$$

The data matrix $\mathcal{X}$ in (19) now takes the form

$$\mathcal{X} = \begin{bmatrix} \boldsymbol{x}_1^1 & \boldsymbol{x}_1^2 & \ldots & \boldsymbol{x}_1^p \\ \vdots & \vdots & & \vdots \\ \boldsymbol{x}_{n-M+1}^1 & \boldsymbol{x}_{n-M+1}^2 & \ldots & \boldsymbol{x}_{n-M+1}^p \end{bmatrix} \tag{22}$$

which is similar to traditional data matrix X in (2) except that now its elements are vectors.

The vector $\boldsymbol{x}_t^s$ in (20) is normally referred to as the *delayed vector* obtained from the time series (x_t^s), $t = 1, \ldots n$ of the field value at grid point s. The new data matrix (22) is now of order $(n - M + 1) \times pM$ which is significantly larger than the original matrix dimension.

We suppose that $\mathcal{X}$ in (22) has been centered and weighted etc. The covariance matrix of (22) is

$$\boldsymbol{\Sigma} = \frac{1}{n - M + 1} \mathcal{X}^T \mathcal{X} = \begin{bmatrix} C_{11} & C_{12} & \ldots & C_{1M} \\ C_{21} & C_{22} & \ldots & C_{2M} \\ \vdots & \vdots & & \vdots \\ C_{M1} & C_{M2} & \ldots & C_{MM} \end{bmatrix} \tag{23}$$

where each C_{ij}, $1 \leq i, j \leq M$ is a lagged covariance matrix between gridpoint i and gridpoint j given by[6]

$$C_{ij} = \frac{1}{n - M + 1} \sum_{k=1}^{n-M+1} \boldsymbol{x}_k^{i\,T} \boldsymbol{x}_k^j \tag{24}$$

[5]also known as *embedding dimension*. This concept is rooted in the theory of dynamical systems.

[6]Other alternatives to compute C_{ij} also exist and they are related to the way the lagged covariance between two time series is computed.

If the elements of the data matrix (22) were random variables and where the covariance matrix is $\mathbf{\Sigma} = E\left(\mathcal{X}^T\mathcal{X}\right)$, then each submatrix $C_{ij} = E\left(\boldsymbol{x}^{iT}\boldsymbol{x}^j\right)$ will exactly take a symmetric Toeplitz form, i.e. with constant diagonals and consequently $\mathbf{\Sigma}$ will be bloc Towplitz. Due to finite sampling, however, C_{ij} is approximately Toeplitz for large values of n (compared to the window length M). This is in general the case when we deal with high frequency data, e.g. daily observations or even monthly averages from climate models. The covariance matrix $\mathbf{\Sigma}$ is therefore a symmetric approximately bloc-Toeplitz for large values of n. Alternative form of the data matrix is provided by writing the state vector (18) in the form

$$\boldsymbol{x}_t = (x_{t1}, \ldots x_{t,p}, x_{t+1,1}, \ldots x_{t+1,p}, \ldots x_{t+M-1,1}, \ldots x_{t+M-1,p}) \tag{25}$$

that is

$$\boldsymbol{x}_t = (\mathbf{x}_t, \mathbf{x}_{t+1}, \ldots, \mathbf{x}_{t+M-1}) \tag{26}$$

where $\mathbf{x}_t$ is the state vector at time t, $t = 1, \ldots n - M + 1$, i.e.

$$\mathbf{x}_t = (x_{t1} \ldots x_{t,p})\,.$$

Hence the matrix (22) now takes the alternative form[7]

$$\mathcal{X}_1 = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_M \\ \vdots & \vdots & & \vdots \\ \mathbf{x}_{n-M+1} & \mathbf{x}_{n-M+2} & \ldots & \mathbf{x}_n \end{bmatrix} \tag{27}$$

This form is exactly equivalent to (22) since it is obtained from (22) by a permutation of the columns as

$$\mathcal{X}_1 = \mathcal{X}P \tag{28}$$

where $P = (p_{ij})$, $i, j = 1, \ldots Mp$, is a permutation matrix[8] given by

$$p_{ij} = \delta_{i,\alpha} \tag{29}$$

where α is a function of j given by $\alpha = rM + \left[\frac{j}{p}\right] + 1$ where $j - 1 \equiv r(p)$, and $[x]$ is the integer part of x.

5.2.3 Computation of EEOFs

EEOFs are the EOFs of the extended data matrix (19), i.e. the eigenvectors of the grand covariance matrix $\mathbf{\Sigma}$ given in (23). They can be obtained directly by computing

[7] used by Weare and Nasstrom 1982

[8] which contains exactly 1 in every line and every column and zeros elswhere. A permutation matrix P is orthogonal, i.e. $PP^T = P^TP = I$.

the eigenvalues/eigenvectors of (23). Alternatively, we can use again the SVD of the grand data matrix $\mathcal{X}$ in (22) in a similar way to (13). This yields

$$\mathcal{X} = V\Theta U^T \tag{30}$$

where $U = (u_{ij}) = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_d)$ represents the matrix of the d extended EOFs or left singular vectors of $\mathcal{X}$, where $d = Mp$ represents the number of new variables, i.e. the number of columns of the grand data matrix. The diagonal matrix Θ contains the singular values $\theta_1, \ldots \theta_d$ of $\mathcal{X}$, and $V = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_d)$ is the matrix of the right singular vectors or extended PC's, i.e. the k'th extended PC is $\boldsymbol{v}_k = (v_k(1), \ldots, v_k(n - M + 1))^T$. These extended EOFs and PCs can be used to filter the data by removing the contribution from nonsignificant components or for reconstruction purposes as detailed below

5.3 Data filtering and oscillation reconstruction

The extended EOFs U can be used as a filter exactly like EOFs. For instance the SVD decomposition (30) yields the expansion of each row $\boldsymbol{x}_t$ of $\mathcal{X}$ in (22)

$$\boldsymbol{x}_t^T = \sum_{k=1}^{d} \theta_k v_k(t) \boldsymbol{u}_k \tag{31}$$

for $t = 1, \ldots n - M + 1$, or in terms of the original variables $\mathbf{x}_t$ as

$$\mathbf{x}_{t+j-1}^T = \sum_{k=1}^{d} \theta_k v_k(t) \mathbf{u}_k^j \tag{32}$$

for $t = 1, \ldots n - M + 1$, and $j = 1, \ldots M$, and where

$$\mathbf{u}_k^j = \left(u_{j,k}, u_{j+M,k}, \ldots, u_{(p-1)M,k}\right)^T.$$

Note that the expression of the vector $\mathbf{u}_k^j$ depends on the form of the data matrix. The one given above corresponds to (22), whereas for the data matrix $\mathcal{X}_1$ (27) we get

$$\mathbf{u}_k^j = \left(u_{(j-1)p+1,k}, u_{(j-1)p+2,k}, \ldots, u_{jp,k}\right)^T$$

Note that when we filter out high ranked EEOFs expression (32) is to be truncated to the required order $d_1 < d$.

The expansion (32) is exact by construction. However, when we truncate it by keeping a smaller number of EEOFs for filtering purposes, e.g. when we reconstruct the field components from a single EEOF or a pair of EEOFs corresponding for example to an oscillation, then the previous expansion does not give a complete picture. This is because (32) truncated to a smaller subset K of EEOFs, e.g.

$$\mathbf{y}_{t+j-1}^T = \sum_{k \text{ in } K} \theta_k v_k(t) \mathbf{u}_k^j \tag{33}$$

where $\mathbf{y}_t = (y_{t,1}, \ldots, y_{t,p})$ is the filtered or reconstructed state space vector, becomes a multivalue function. For example, for $t = 1$ and $j = 2$, we get one value of $y_{t,1}$ and for $t = 2$ and $j = 1$ we get another value of $y_{t,1}$. Note that this is due to the fact that the EEOFs have time lagged components. To get a unique recontructed value we simply take the average of those multiple values. The number of multiple values depends[9] on the value of time $t = 1, \ldots n$. The reconstructed variables using a subset K of EEOFs are then easily obtained from (33) by

$$\mathbf{y}_t^T = \begin{array}{l} \frac{1}{t}\sum_{j=1}^{t}\sum_K \theta_k v_k(t-j+1)\mathbf{u}_k^j \quad \text{for} \quad 1 \le t \le M-1 \\ \frac{1}{M}\sum_{j=1}^{M}\sum_K \theta_k v_k(t-j+1)\mathbf{u}_k^j \quad \text{for} \quad M \le t \le n-M+1 \\ \frac{1}{n-t+1}\sum_{j=t-n+M}^{M}\sum_K \theta_k v_k(t-j+1)\mathbf{u}_k^j \quad \text{for} \quad n-M+2 \le t \le n \end{array} \tag{34}$$

Note that these constructions can also be obtained in a least square sense (see, e.g. Vautard et al. 1992, and Ghil et al. 2001). the reconstructed components can also be restricted to any subset of the eigen elements of the grand data matrix (22) or similarly the grand covariance matrix $\mathbf{\Sigma}$. For example to reconstruct the time series associated with an oscillatory eigen elements, i.e. a pair of degenerate eigenvalues, the subset K in the sum (34) is limited to that pair.

The reconstructed multivariate time series $\mathbf{y}_t$, $t = 1, \ldots n$, can represent the reconstructed (or filtered) values of the original field at the original p grid points. In general, however, the number of grid points is too large to warrant an eigen-decomposition of the grand data or covariance matrix. In this case a dimension reduction of the data is first applied by using the p_0, say, leading PCs then apply a MSSA to these retained PCs. In this case the dimension of $\mathcal{X}$ becomes $(n-M+1) \times Mp_0$, which may be made considerably smaller than the original dimension. To get the reconstructed space-time field one then use the reconstructed PCs in conjunction with the p_0 leading EOFs.

5.4 Application to outgoing long wave radiation

The outgoing long wave radiation (OLR) comes from NCAR/NCEP reanalyses over the tropical region from 30^{o}S to 30^{o}N. This analysis focuses on a 5-year period of daily data from 1 Jan 1996 to 31 Dec 2000. Figure 5 displays the OLR field on 25 Dec 1996. Note the low-value regions particularly over the warm pool, an area of large and intensive convection, and to a lesser extent over the Amazonian and tropical African

[9]These numbers can obtained by constructing a $M \times n$ array $A = (a_{jt})$ with entries $a_{jt} = t-j+1$. Next all entries that are nonpositive or greater than $n - M + 1$ are to be equated to zero. Then for each time t take all the indices j with positive entries.

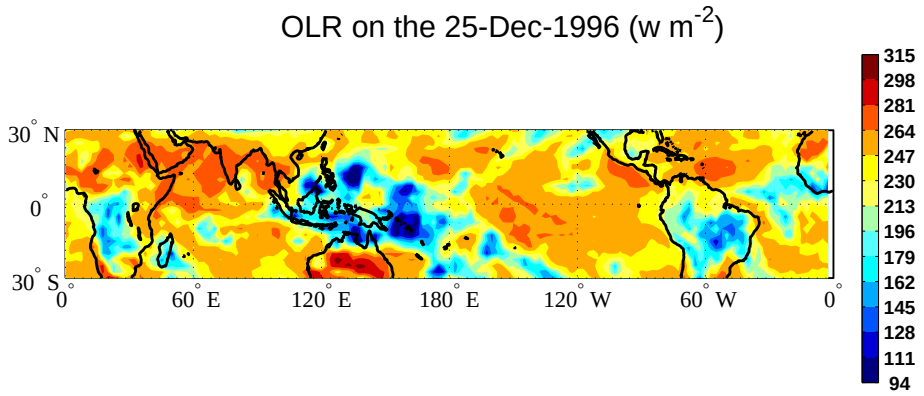

Figure 5. The distribution of OLR (w/m^2) on the 25-Dec-1996 over the tropics.

regions. The OLR data is not very homogeneous, and have a quite complicated variability as well as seasonality. To illustrate this complication, figure 6 shows the OLR time series at four different locations. A clear indication of nonstationarity of the times series is obvious. Note for example the high values at the equator and 122.5^{o}E in winter 97/98. This period corresponds to a strong El-Nino event where the convection shifts eastward to the mid-Pacific allowing more long wave radiation to be lost to space over the maritime continent. Note also the strong seasonal component at 30^{o}N.

Before applying EEOF analysis, the data were first subjected to an EOF analysis. The leading 10 EOFs/PCs of the anomaly field with respect to the long term average (climatology) were retained for the analysis. Figure 7 shows the leading EOF mode. This pattern explains about 15% of the total variability and is associated with the seasonal cycle. Fig. 7 clearly indicates that the seasonal cycle is mostly explained by the Inter Tropical Convergence Zone (ITCZ) and some monsoonal activities.

The first 10 EOFs/PCs used for the EEOF/MSSA analysis explain altogether about 32% of the total variability. A window length corresponding to $M = 80$ days is used to construct the grand data matrix (22). This choice was motivated by the desire to capture the Madden-Julian Oscillation (MJO) identified first by Madden and Julian (1971). The extended (grand) data matrix is then factorised using SVD. Figure 8 shows the spectrum of the grand covariance matrix $\mathbf{\Sigma}$ (23), i.e. the square of the singular values of the grand data matrix (22). The standard error are obtained using (12) with a heuristic effective sample size of 116 corresponding to an uncorrelation lag of 15 days. This is just a rough and heuristic estimation because the observation period is not long enough.

The first two eigenvalues of the spectrum (Fig. 8) correspond to the oscillation of the seasonal cycle. They do not look nearly equal and well separated from the rest

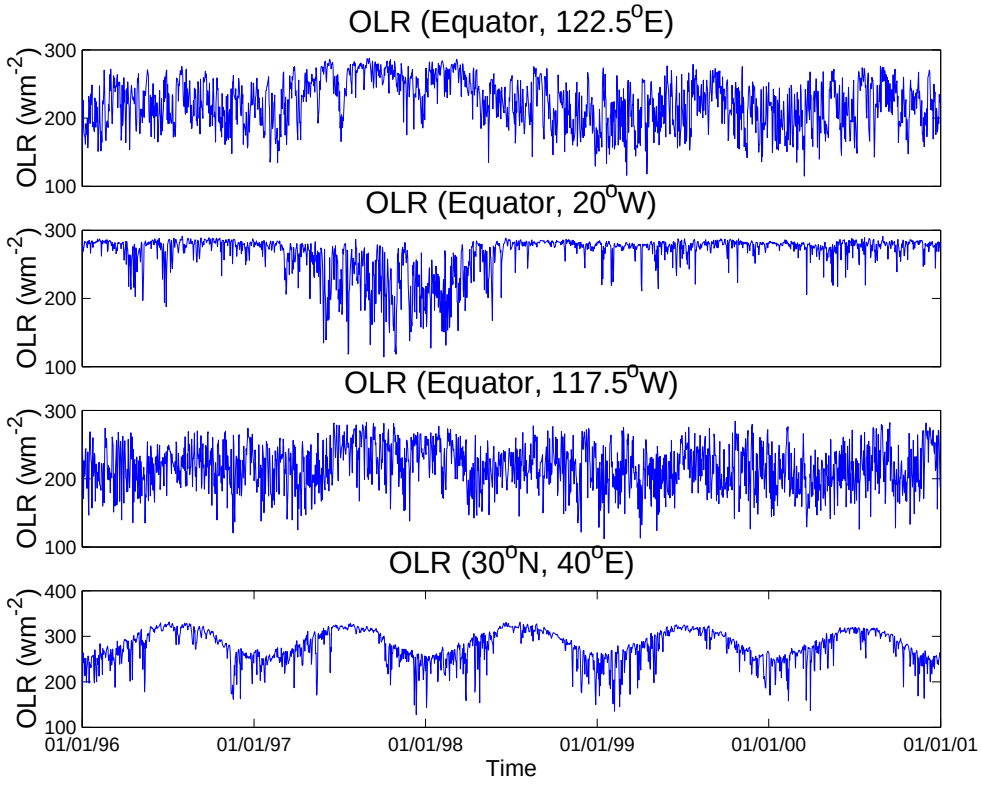

Figure 6. Time series of OLR ($\mathrm{w}m^{-2}$) at four different locations.

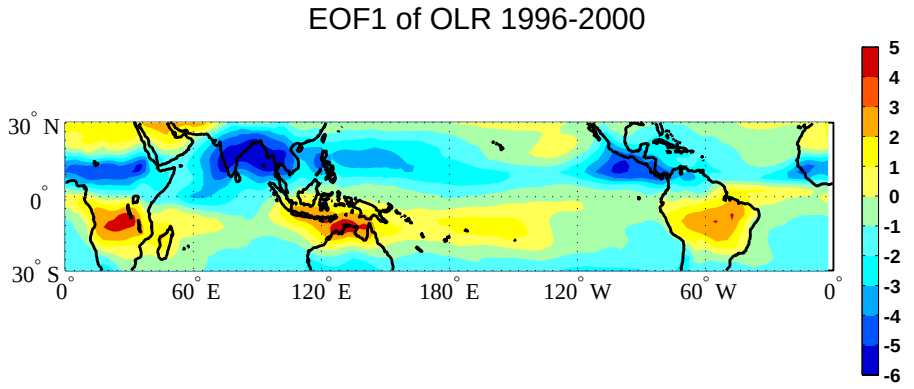

Figure 7. The leading EOF of OLR anomalies. Units are arbitrary.

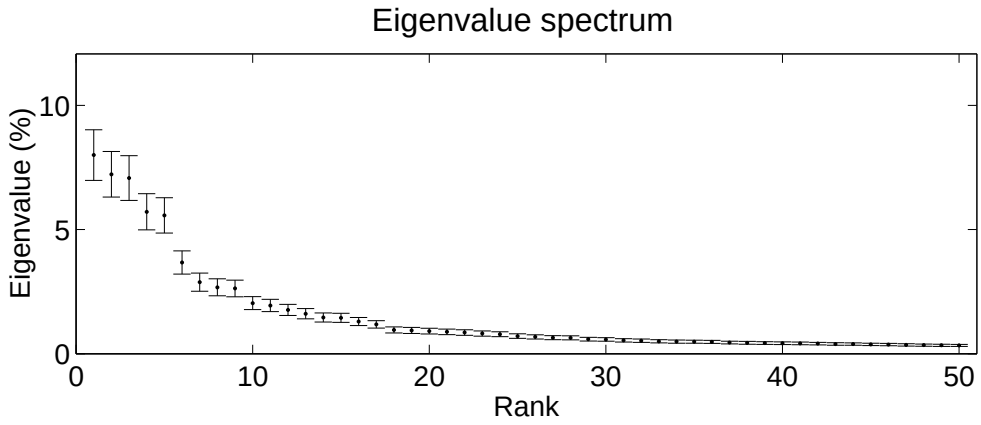

Figure 8. The spectrum of the grand covariance matrix (23). Standard errors are derived from (12) using an effective sample size of ..

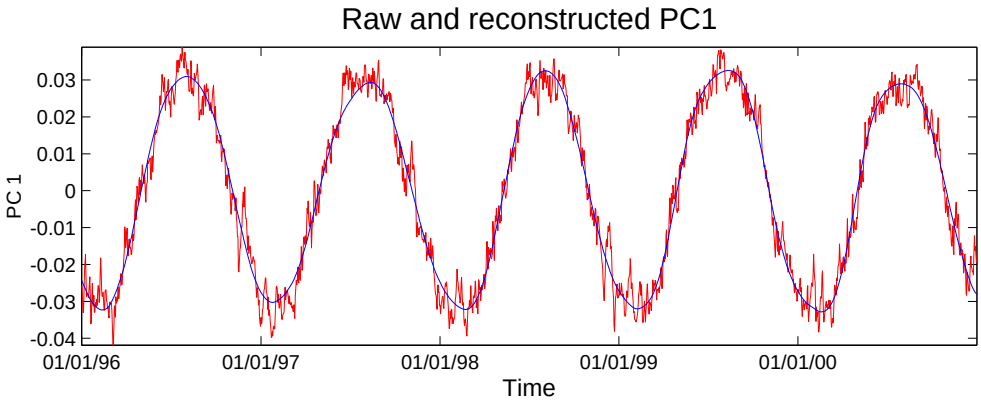

Figure 9. The time series of the raw and reconstructed PC1.

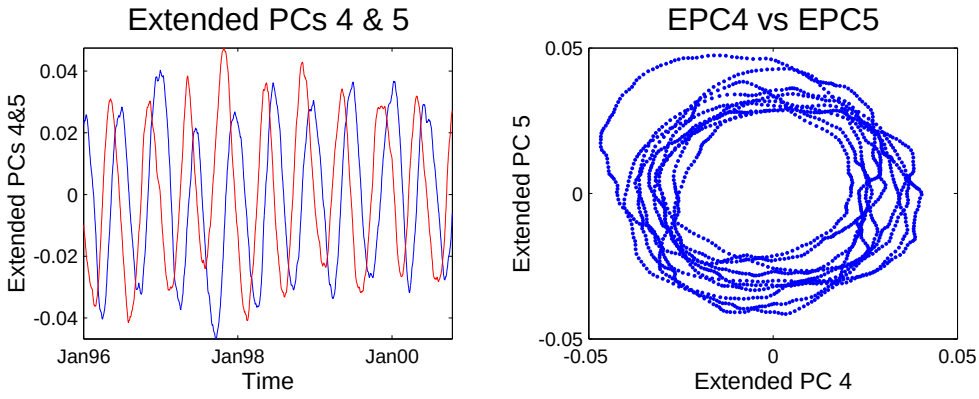

Figure 10. Time plot of EPC4 and EPC5 (left pannel) and phase diagram of EPC4 versus EPC5 (right pannel).

of the spectrum, however, and this is due to the choice of the window length, which is much smaller than the seasonal cycle. Despite this, the first two extended PCs (EPCs) show a pair of sine waves perfectly in quadrature, and represent the seasonal cycle. Figure 9 shows in fact the raw PC1 along with the reconstructed or smoothed PC1. The reconstruction is based on (34) using the leading 10 EPCs.

Beside the annual cycle, the (degenerate) fourth and fifth eigenvalues[10] constitute also another oscillatory pair corresponding to the semi-annual cycle. The left panel of Fig. 10 shows a time plot of EPC4 and EPC5 whereas the phase diagram of EPC4 versus EPC5 is shown in the right pannel of Fig. 10. The figure clearly shows the semi-annual oscillation (SAO) in OLR. The phase diagram (right pannel) also shows the SAO with slight irregularities due to interannual variability. The next oscillatory pair corresponds to the 8'th and 9'th eigenvalues (Fig. 8). This pair of (nearly) equal eigenvalues correspond to the familiar MJO, and corresponds to a period of about 50 days.

The MJO is well studied and well documented in the literature since it was first identified by Madden-Julian (1971) using spectral techniques. It constitutes a well established dominant mode of intraseasonal tropical variability. It is an eastward propagating planetary scale wave of tropical convective anomalies. The oscillation has a quite broad band with period between about 40 and 60 days (Knutson and Weickmann 1987; Hendon and Salby 1994; Madden and Julian 1994). It has been identified from various fields such as zonal and divergent wind, sea level pressure, and OLR in the tropics (Madden and Julian 1972; Kiladis and Weickmann 1992). Hendon and Salby (1994) find that the MJO is composed of a forced response asso-

[10]the degeneracy can be seen by using the rule of thumb (12) but without serial correlation.

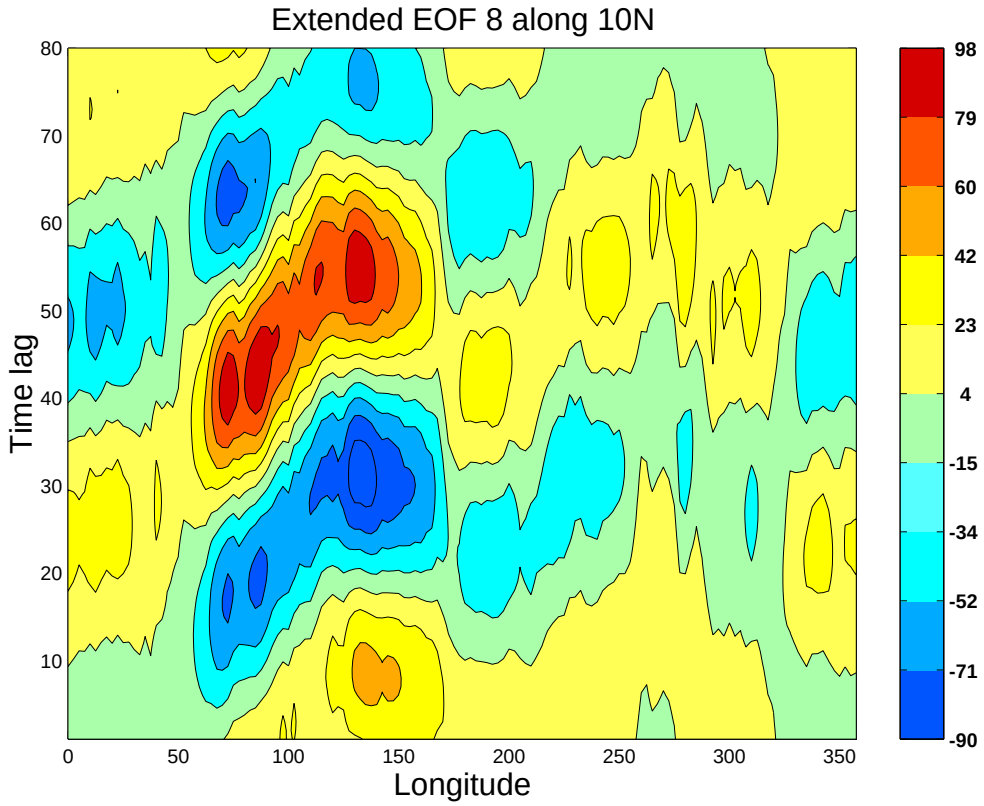

Figure 11. Extended EOF 8 along 10^oN as a function of time lag. Units arbitrary.

ciated with the convective part, obtained, e.g. from OLR, and a radiating response, associated with the circulation part, identified, e.g. from the zonal wind. The circulation anomaly associated with the MJO tends to propagate along the tropics whereas the convective anomaly tends to be confined to the eastern hemisphere (Hendon and Salby 1994).

Various mechanisms have been proposed such as triggering by equatorial Kelvin waves (Matthews 2000) or as a nonlinear interaction in the frequency domain between surface and boundary layer moisture fluxes (Krishnamurti 2003). Hendon and Salby (1994) find that the forced response is an eastward propagating coupled Rossby-Kelvin wave at about 5 m s^{-1} whereas the radiating response appears as a Kelvin wave propagating faster, at about twice the previous speed, away from the convective anomaly. They also suggest that frictional wave−CISK plays a key role in generating the MJO.

Figure 11 shows extended EOF 8 along 10^oN as a function of time lag. This sort of diagram where the space and time axes are shown is known as hovmoller diagram. Figure 11 shows clearly the eastward moving oscillation with an average phase speed around 7^o/day, making the wave travel from west to east in roughly 50 days, i.e.

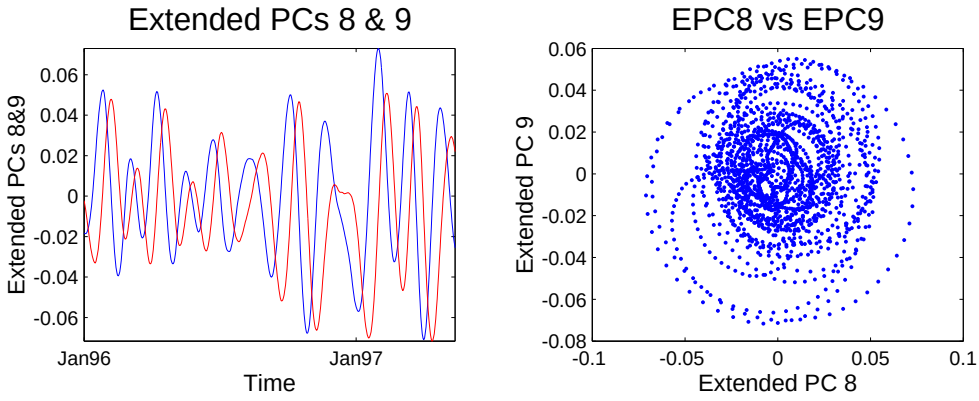

Figure 12. Extended PCs 8 and 9 (left pannel) and phase diagram of EPC8 versus EPC9 (right pannel). arbitrary.

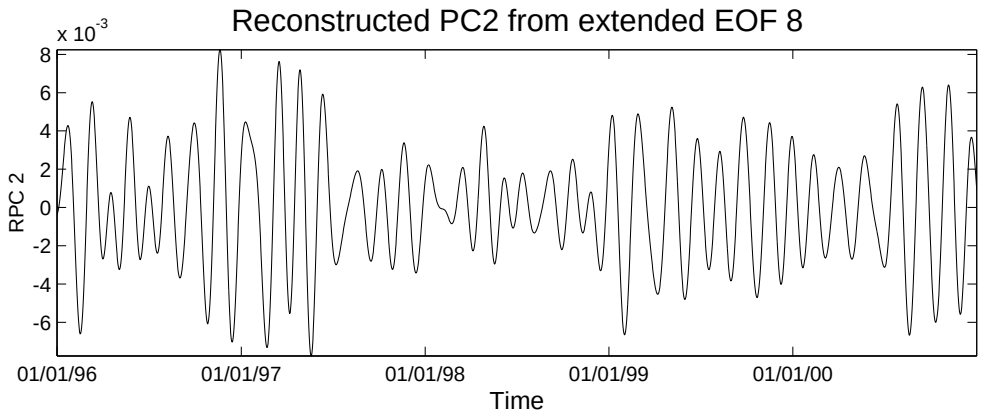

Figure 13. Reconstructed PC2 using the extended EOF/PC 8.

approximately 9 m s^{-1} as observed in Knutson and Weickmann (1987). Figure 12 shows the time plot of the extended PCs 8 and 9 along with the phase diagram of EPC 8 versus EPC 9. The EPCs are oscillating in quadrature, with a period of about 50 days and with varying intensity. The power spectrum of EPC8 (not shown) peaks around 50 days with a slight broad band structure.

The MJO can be found in the reconstructed PCs using the extended EOFs/PCs. Figure 13, for example, shows a time plot of the reconstructed PC2 using the 8'th EEOF/EPC. The figure shows clearly the oscillation with varying amplitude. Note in particular the strong oscillation during winter 97 versus the weak oscillation from winter to early autumn 98. The same behaviour is also observed in the phase diagram in the RPC 6 versus RPC 7 (Fig. 14).

Since EOF analysis does not focus on a particular frequency band, one would expect MJO to project onto more than one PC. In fact, MJO is found to project onto

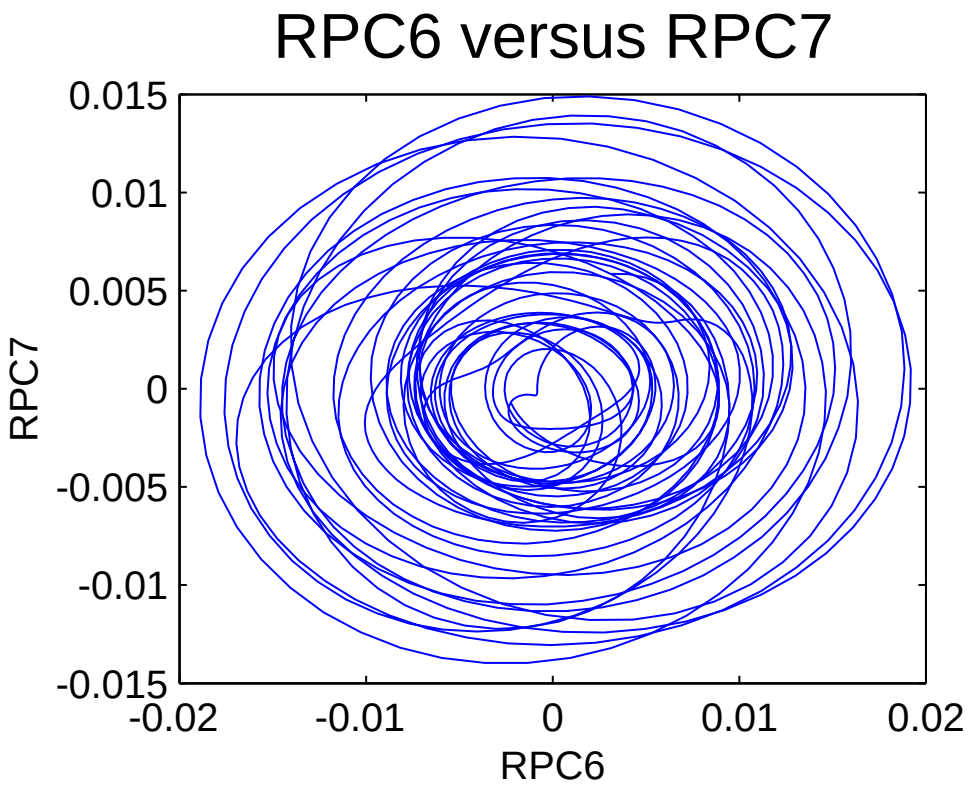

Figure 14. Phase diagram of the reconstructed PC 6 versus the reconconstructed PC 7.

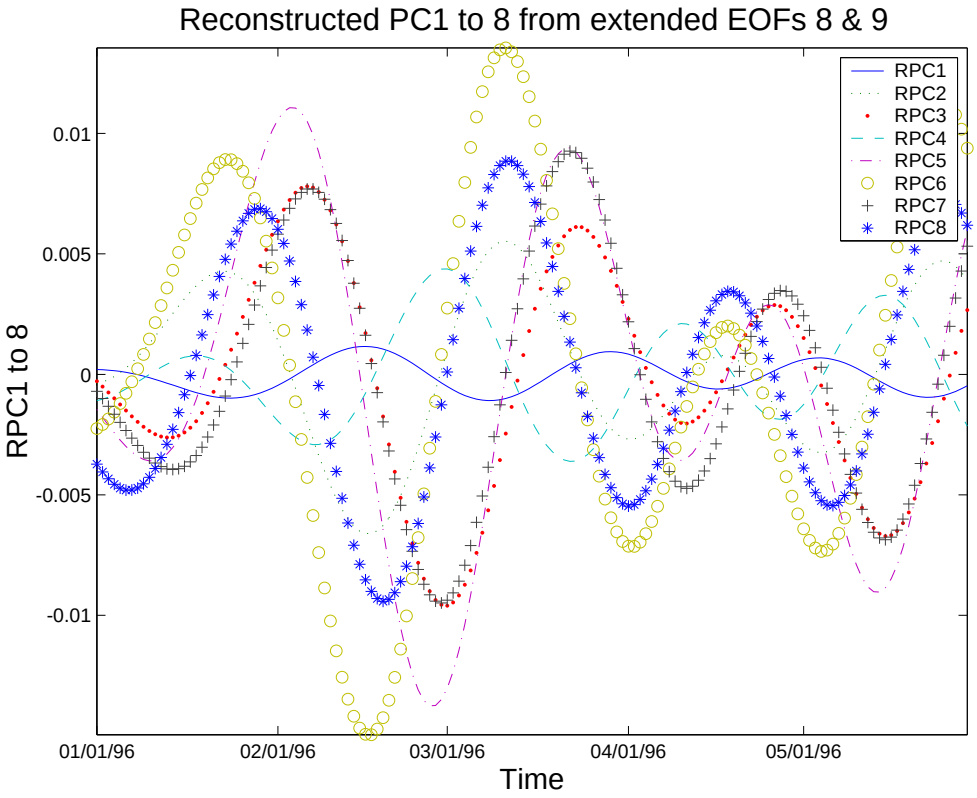

Figure 15. Reconstructed PCs 1 to 8 using the extended EOFs/PCs 8 and 9.

various PCs but with differing energies (amplitudes). For instance SSA analysis of single PCs reveals that MJO is mostly present in higher PCs, e.g. PCs 6 and 7. This observation is also revealed from analysing the reconstructed PC's using the extended EOFs/PCs 8 and 9 associated with the oscillation. This is well illustrated in figure 15, which shows the first 8 reconstructed PCs. Reconstructed PCs 5 to 8 are the most energetic components representing MJO. Note in particular the weak projection of MJO onto the first PC, which represents only the seasonal cycle.

The reconstructed or filtered PCs allow a reconstruction of the original space-time OLR field. Figure 16 shows a hovmoller diagram of the reconstructed OLR field, at 10^oN using the MJO-related extended EOFs/PCs 8 and 9. The reconstructed PC's 1 to 8 (Fig. 15) are used in conjunction with the EOFs as in (11) to get the MJO reconstructed field. It is clear from figure 16 that MJO gets triggered at around 25-30^oE over the African jet region. It reaches its mature stage over the Indian ocean in the Bay of Bengal and starts to decay thereafter. The MJO gets particularly damped near 150^oE over the convective region in the warm pool. Figure 16 shows also the dispersive nature of the MJO with a stronger phase speed during the growth phase compared to that observed during the decay phase.

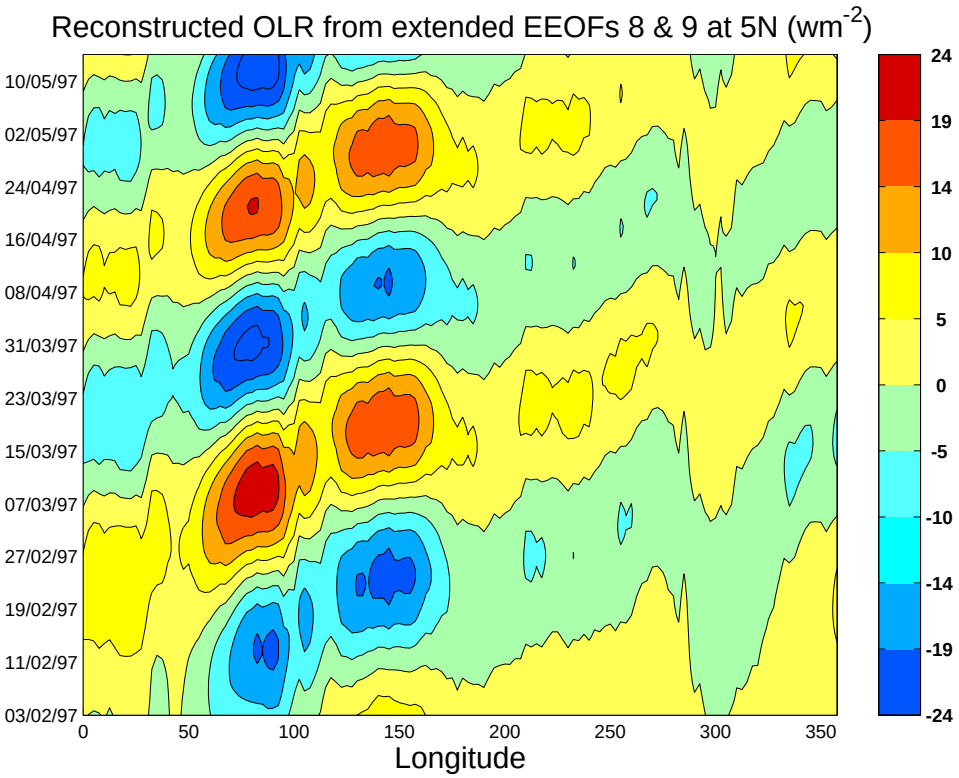

Figure 16. Reconstructed OLR field using the reconstructed EOFs/PCs 1 to 8 shown in Fig. 15.

6 Summary and conclusion

We have explained in this primer how to derive and compute EOFs for climate gridded data. We pointed out particularly the meaning of EOFs and their physical interpretability. We have also highlighted the major drawbacks of the method, namely orthogonality of EOFs and uncorrelation of PCs. In addition, the truncation order for signal/background noise separation remains arbitrary. We have illustrated the method by applying it to the winter monthly SLP. We obtained in particular the familiar Arctic Oscillation pattern as leading EOF and the Pacific pattern as second EOF. The first EOF seems nondegenerate but the remainder of the spectrum looks continuous.

Rotated EOFs are then presented as a way to overcome some of the previous drawbacks related to orthogonality and uncorrelation. REOFs also help overcome the problem of degenerate eigenspectrum. They attempt to maximise a simplicity criteria by rotating a fixed number of EOFs. We have also illustrated the method with the winter monthly SLP. The method yields the NAO as the leading REOF and the Pacific pattern as second REOF. We have also highlighted the dependence of REOFs on the number of retained EOFs and also the rotation criteria. For example, when the number of retained EOFs increases the REOFs become more and more localised.

Extended EOFs (or MSSA) are presented as a way to overcome some of the shortcomings of EOFs, namely the use of only spatial correlation. MSSA makes use of spatial as well as temporal correlation by extending the familiar state vector by including explicitely the time information. The method can be used as a tool to filter the data, isolate a trend, or even sparate an oscillatory component burried in the noisy data. We have illustrated the approach with 5 years of daily OLR data. We have in particular identified the seasonal cycle and the semi-annual oscillation. The MJO was also identified from the spectrum of the extended data (or trajectory) matrix. various characteristics of MJO have also been identified such as the most active region of growth and decay phase, approximate period, and phase speed.

Acknowledgement.

This work was supported by the Centre for Global Atmospheric Modeling at the Department of Meteorology, the University of Reading. I would like to thank particularly Prof J. Slingo for her interest in and encouragement of this work, Dr D. B. Stephenson for his encouragement and his constructive critics. Particular thanks also go to Dr C. Ferro for thouroughly reading the paper and providing positive comments, and Prof I. Jolliffe for his constructive comments, Dr S. Pezzulli for his encouragement.

Thanks also to Dr P. Inness for providing the OLR data.

References

Angström, A., 1935: *Geografiska Annaler*, **17,** 242-258.

Bjerknes, J., 1969: Atmospheric teleconnections from the equatorial Pacific. *Mon. Wea. Rev.,* **97**, 163-172.

Broomhead, D. S., and G. P. King, 1986a: Extracting qualitative dynamics from experimental data. *Physica D*, **20**, 217-236.

Broomhead, D. S., and G. P. King, 1986b: On the qualitative analysis of experimental dynamical systems. *Nonlinear Phenomena and Chaos*, S. Sarkar, Ed., Adam Hilger, 113-144.

Caroll, J. B., 1953: An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, **18,** 23-38.

Craddock, J. M., 1973: Problems and prospects for eigenvector analysis in meteorology. *The statistician*, **22**, 133-145.

Dommenget, D., and M. Latif, 2002: A cautionary note on the interpretation of EOFs. *J. Climate*, **15,** 216-225.

Fraedrick, K., 1986: Estimating the dimensions of weather and climate attractors. *J. Atmos. Sci.,* **43,** 419-432.

Fukuoka, A., 1951: A study of 10-day forecast (A synthetic report). *The Geophysical Magazine, Tokyo*, Vol. XXII, 177-218.

Hendon, H. H., and M. L. Salby, 1994: The life cycle of the Madden-Julian oscillation. *J. Atmos. Sci.,* **51,** 2225-2237.

Hotelling, H., 1933: Analysis of a complex of statistical variables into principal components. *J. Educ. Psych,*, **24,** 417-520.

Jolliffe, I. T., 2002: *Principal Component Analysis*. Springer-Verlag, 2nd Edition, New York.

Kiladis, G. N., and K. M. Weickmann, 1992: Circulation anomalies associated with tropical convection during northern winter. *Mon. Wea. Rev.,* **120,** 1900-1923.

Kimoto, M., M. Ghil, and K. C. Mo, 1991: Spatial structure of the extratropical 40-day oscillation . *Proc. 8'th Conf. Atmos. Oceanic waves and Stability*, Amer. Meteor. Soc., Boston, 115-116.

Knutson, T. R., and K. M. Weickmann, 1987: 30-60 day atmospheric oscillations: composite life cycles of convection and circulation anomalies. *Mon. Wea. Rev.,* **115,** 1407-1436.

Krishnamurthi, T. N., D. R., Chakraborty, N., Cubucku, L., Stefanova, and T. S. V., Vijaya Kumar, 2003: A mechanism of the Madden-Julian oscillation based on interactions in the frequency domain. *Q. J.*

R. Meteorol. Soc., **129,** 2559-2590.

Kutzbach, J. E., 1967: Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over North America. *J. Appl. Meteor.*, **6,** 791-802.

Linz, P., and R. L. C. Wang, 2003: *Exploring Numerical Methods: An Introduction to Scientific Computing Using MATLAB.* Jones and Bartlett Publishers, Sudbury, Massachusetts.

Loève, M., 1978: *Probability theory*, Vol II, 4'th ed., Springer-Verlag, 413 pp.

Lorenz, E. N., 1970: Climate change as a mathematical problem. *J. Appl. Meteor.,* **9,** 325-329.

Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. Technical report, Statistical Forecast Project Report 1, Dept. of Meteor., MIT, 1956. 49pp.

Madden, R. A., and P. R. Julian, 1971: Detection of a 40-50 day oscillation in the zonal wind in the tropical pacific. *J. Atmos. Sci.,* **28,** 702-708.

Madden, R. A., and P. R. Julian, 1972: Description of global-scale circulation cells in the tropics with a 40-50 day period. *J. Atmos. Sci.,* **29,** 1109-1123.

Madden, R. A., and P. R. Julian, 1994: Observations of the 40-50-day tropical oscillation−A review. *Mon. Wea. Rev,* **122,** 814-837.

Matthews, A. J., 2000: Propagation mechanisms for the Madden-Julian oscillation. Q. J. R. Meteorol. Soc., **126,** 2637-2651.

North, G. R., T. L., Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Weather Rev.,* **110,** 699-706.

Obukhov, A. M., 1960: The statistically orthogonal expansion of empirical functions. *Bull. Acad. Sci. USSR Geophys. Ser.* (English Transl.), 288-291.

Obukhov, A.M., 1947: Statistically homogeneous fields on a sphere. *Usp. Mat. Navk.,* **2,** 196-198.

Plaut, G., and R. Vautard, 1994: Spells of low-frequency oscillations and weather regimes in the northern hemisphere. *J. Atmos. sci.*, **51**, 210-236.

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

Simmons, A. J., G. W. Branstator, and J. M. Wallace, 1983: Barotropic

wave propagation, instability and atmospheric teleconnection paaterns. *J. Atmos. Sci.,* **40,** 1363-1392.

von Storch, H., 1995: Spatial patterns: EOFs and CCA. In H. von Storch and A. Navarra, editors, *Analysis of Climate Variability: Application of Statistical Techniques*, Springer Verlag, 227-257.

von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate research,* Cambridge University Press, Cambridge.

Thompson, D. W. J., and J. M. Wallace, 1998: The Arctic Oscillation signature in wintertime signature in wintertime geopotential height and temperature fields. *Geophys. Res. Lett,* **25,** 1297-1300.

Thompson, D. W. J., and J. M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate,* **13,** 1000-1016.

Thurstone, L. L., 1940: Current issues in factor analysis. *Psychological Bulletin*, **37,** 189-236.

Thurstone, L. L., 1947: *Multiple Factor Analysis*, The University of Chicago Press, Chicago.

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

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Weath. Rev.*, **109,** 784-812.

Weare, B. C., and J. S. Nasstrom, 1982: Examples of extended empirical orthogonal function analysis. *Mon. Weath. Rev.*, **110**, 481-485.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, San Diego.