The DOI for this manuscript is doi: 10.1175/2008JCLI2147.1

Spatial weighting and iterative projection methods for EOFs

Mark P. Baldwin[1], David. B. Stephenson[2], Ian T. Jolliffe[2]

Submitted to the Journal of Climate, June 2007

Revised, June, 2008

[1] NorthWest Research Associates, Seattle, WA, USA. mark@nwra.com.

[2] School of Engineering, Computing, and Mathematics, University of Exeter, UK.

Abstract

Often there is a need to consider spatial weighting in methods for finding spatial patterns in climate data. In this paper we focus on techniques that maximize variance, such as Empirical Orthogonal Functions (EOFs). We introduce a weighting matrix into a generalized framework for dealing with spatial weighting. One basic principal in the design of the weighting matrix is that the resulting spatial patterns are independent of the grid used to represent the data. A weighting matrix can also be used for other purposes, such as to compensate for the neglect of unrepresented sub-grid scale variance or, in the form of a prewhitening filter, to maximize the signal-to-noise ratio of EOFs. Our methodology is applicable to other types of climate pattern analysis, such as extended EOF analysis and Maximum Covariance Analysis. The increasing availability of large datasets of 3-dimensional gridded variables (e.g., reanalysis products and model output) raises special issues for data reduction methods such as EOFs. Fast, memory-efficient methods are required in order to extract leading EOFs from such large data sets. This study proposes one such approach based on a simple iteration of successive projections of the data onto time series and spatial maps. We also demonstrate that spatial weighting can be combined with the iterative methods. Throughout the paper, we use multivariate statistics notation that can be implemented as simple matrix commands in high-level computing languages.

1. Introduction

The analysis of spatial patterns in geophysical data is performed with a wide variety of techniques, including methods based on Empirical Orthogonal Functions (EOFs; e.g., von Storch and Zwiers, 1999; Jolliffe, 2002; Bretherton, 2003; Hannachi et al., 2006; Hannachi et al., 2007; van den Dool, 2007), Empirical Normal Modes (Brunet, 1994; Brunet and Vautard, 1996), methods that find coupled patterns between data sets (e.g., Maximum Covariance Analysis, Canonical Correlation Analysis, Combined Principal Component Analysis; see von Storch and Zwiers, 1999), methods that find similar patterns, such as cluster analysis (e.g., Cheng and Wallace, 1993; Wallace et al., 1993a) and self-organized maps (Hewitson and Crane, 2002; Reusch et al., 2007). With any of these methods, the use of spatial weighting may be an important issue for several reasons.

One reason to use spatial weighting is that spatial patterns should be invariant to how one chooses the grid locations, since one is aiming to find properties of the continuous spatial field. e.g., "intrinsic EOFs" (North et al. 1982; Stephenson, 1997). Spatial weighting can be used to compensate for unequal distribution of grid points. Weighting may be desirable for other purposes, such as to emphasize (or mask) certain spatial regions, to account for variations in error covariances, to calculate patterns with more than one variable measured across the spatial domain (e.g., extended EOF analysis), or to equalize the variance at every grid point (e.g., EOFs based on the correlation matrix instead of the covariance matrix). Weighting may also be used to compensate for small-scale variance that is not represented by the gridded data matrix. It may also necessary to apply weighting to find EOFs of quantities such as zonally-averaged angular momentum, derived from the zonal-mean zonal wind field (e.g., Baldwin and Tung, 1994). Weighting may also take the form of a prewhitening filter (Allen and Smith, 1997; Venzke et al., 1999; Chang et al., 2000) in which the data matrix is premultiplied by a filter (weighting matrix) that is determined by the noise covariance matrix.

In this paper we adopt a general approach to spatial weighting of geophysical fields. We use EOF analysis as an example to introduce a generalized weighting matrix. We show that it is possible to include the weighting matrix in the EOF calculation, without premultiplying the data by the square roots of the weights, as was done by North et al. (1982). Our results are applicable to a wide variety of similar techniques—in general any techniques that partition variance.

It is often desirable to perform EOF analysis on large, possibly 3-D, spatio-temporal data sets (e.g., Hawkins and Sutton, 2007). Standard EOF techniques can become impractical for large data matrices. Efficient techniques such as the power method can be used to find the leading EOFs (Golub and van Loan, 1983; Jolliffe, 2002 for a review; van den Dool, 2007), but typically involve the explicit calculation of the whole covariance matrix, which can make even these techniques impractical. It becomes necessary to have EOF methods that can rapidly find the leading EOFs for such data sets, while allowing for arbitrary spatial weighting.

This paper addresses these issues by a) proposing a simple iterative scheme suitable for finding EOFs in large data sets, and b) mathematically formulating the EOF problem for arbitrary weighting schemes. The iterative method is applicable to any of the EOF-based techniques and uses the data matrix directly. We demonstrate that the convergence properties of the iterative method are the same as those of the power method (Golub and van Loan, 1983).

The outline of the paper is as follows. Section 2 introduces the nomenclature and fundamentals of EOFs in standard matrix notation. Section 3 discusses spatial weighting in EOF analysis; In Section 3a we introduce a weighting matrix, and in Section 3b we discuss how to deal with sub-gridscale variance. Section 3c presents a derivation of a generalized weighting metric, which can be easily transformed away by a change of variables involving multiplying the data by the square root of the weighting metric. Section 4a discusses in general how to "project" data onto time series and spatial patterns. We then show, in Section 4b, how an iterative projection method can be combined with spatial weighting to efficiently calculate the leading EOF. In section 5 we provide examples of including weighting matrices in EOF analysis.

2. Nomenclature and approach to EOF analysis

We begin by considering a field, such as pressure or temperature, that is a continuous function of space and time. We assume that the continuous field is approximated by sampling at regular time intervals onto some spatial grid, which may or may not be regular. The data are organized in an $(n \times p)$ matrix, $X = [x_1, x_2, \ldots, x_n]^T$, containing n observations of a variable x_t defined at p spatial points. The p spatial points could be embedded in one, two or three spatial dimensions. The n observations could be of the same field made at different times, or n observations of different images (e.g. images of human faces – Craw and Cameron, 1992). For the purposes of this paper we will assume, without loss of generality, that we are dealing with p spatial observations at n regularly spaced times.

We also assume the variables in X (the time series, or columns) are anomalies that have been obtained by centering the original variables by removing the respective time means at each location. The means of each time series (i.e., the columns of X) are therefore identically zero. Because the variables are centered, the $(p \times p)$ sample covariance matrix is given by the matrix product $S = \frac{1}{n-1} X^T X$. Here it is conventional to use n-1 (instead of n) when the covariance is computed from a sample. The *total variance* is the sum of the variances at each of the grid points (the diagonal elements of S) and so is given by the trace of S, $Tr(S)$.

EOF analysis decomposes a data matrix into a series of data-based orthogonal functions, which are defined so that a minimal number of EOFs are needed to reconstruct the variation within the original data matrix. EOF analysis can be understood in terms of spatial patterns (EOFs) and their associated time series (often called Principal Components). The data matrix can be written as a sum of the products of the EOFs $\{e_1, e_2, \ldots\}$ and their associated time series $\{y_1, y_2, \ldots\}$:

$$X = \sum_{i=1}^{r} y_i e_i^T \tag{2.1}$$

where r is the rank of X, which is never greater than the minimum of n and p. The EOFs $\{e_1, e_2, \ldots\}$ are chosen so as to be orthonormal ($e_i^T e_j = 1$ for $i = j$ and 0 otherwise) and successively maximize the variance in the corresponding time series $\{y_1, y_2, \ldots\}$. $e_1^T S e_1$ is maximal and each successive EOF component accounts for the greatest possible fraction of the remaining variance. The EOF time series can be shown to be uncorrelated with one another, but this does not necessarily mean that they are independent. For example, a pair of EOFs may describe a propagating pattern in a data matrix (e.g., the quasi-biennial oscillation, as shown by Wallace et al., 1993b).

3. Spatial weighting in EOF analysis

One common reason to weight in EOF analysis is to compensate for grid spacing in the data matrix, for example to compensate for converging meridians on a latitude-longitude grid. But there are several other reasons to consider weights in EOF analysis. It may be necessary to create a metric of some integral quantity of interest, such as forming angular momentum from zonal wind data or weighting atmospheric data in the vertical by inverse pressure or by height. It may be desirable to apply different weights to each variable when several variables are included in the analysis. In some cases one may wish to emphasize or de-emphasize certain geographic regions or data levels. Weighting may be used to correct for spatially inhomogeneous data variances or error covariances. EOF analysis is often done using correlations rather than covariances, which corresponds to weighting anomalies by the reciprocal of their respective standard deviations.

As discussed in North et al. (1982), there is typically a finite number of sometimes irregularly spaced grid points in a domain. We wish to find what North et al. termed the "intrinsic EOFs," which are independent of the data grid. The concept of intrinsic patterns, independent of the grid, can be applied more broadly to any type of climate pattern analysis. To do this we must find an appropriate weighting that is invariant to the choice of grid spacing.

We will use EOF analysis as an example, but the resulting weighting methods are broadly applicable to all variants of EOF methods [e.g., Empirical Orthogonal Teleconnections (van den Dool et al., 2000)]. This weighting method is also applicable to Maximum Covariance Analysis (MCA) because MCA with identical input fields finds the EOFs. For any of these methods, the weighting technique must produce the same patterns and time series as would result from regridding the data onto an equal area grid.

Despite all these reasons, the subject of how to weight EOFs has received little formal attention in the published scientific literature. This section will attempt to address this deficit by showing how EOF analysis can be generalized to include arbitrary weighting metrics.

3a. Weighting metric

The generalization of the EOF problem to continuous (Karhunen-Loève) functions was discussed by North et al. (1982). They reasoned that EOFs derived from an approximation to the continuous field, on a finite set of grid points, should approximate those based on the continuous functions, and should therefore be independent of the particular grid chosen. Our approach is similar in that we begin by considering a continuous anomaly field.

It is important to note that EOF methods partition the total variance of the data matrix, which is a sum of *squared* anomalies from the mean. For the purpose of introducing a weighting matrix, it is sufficient to consider a continuous anomaly field, x, at a single time, such as sea-level pressure over the Northern Hemisphere. The area integral of x^2 would correspond to a summation of squared grid point values. If x is approximated on a series of p grid points, then the area integral of the squared continuous anomaly field x^2 can be approximated by the quadratic sum of anomaly values at the grid points i. If the grid is uniform, and therefore weighting is not needed,

$$\int_A x^2 da \approx \sum_{i=1}^{p} x_i^2 = x^T x \qquad (3.1)$$

To accommodate the most general types of weighting, we introduce the $(p \times p)$ weighting matrix W.

$$\int_A x^2 da \approx \sum_{i=1}^{p} \sum_{j=1}^{p} x_i W_{ij} x_j = x^T W x \qquad (3.2)$$

We define the weighting matrix in a general way so that it can be used to area weight the summation, or to weight or filter EOFs in other circumstances (described below). It is a generalization of the area weighting concept discussed in North et al. (1982), who defined a p-vector of weights proportional to the grid box area.

W can always be written symmetrically $W_{ij} = W_{ji}$ without any loss of generality. Since the integral of a squared quantity is non-negative, it is desirable that $x^T W x \geq 0$ and so the weighting matrix W then defines a metric $\| x \|_W^2 = x^T W x$ in the p-dimensional space of grid point variables (Strang, 2003).

It is common practice to assume that the area integral can be approximated by the simple sum of squared anomalies at each grid point—in other words, the mid-point rule for numerical integration. In the simplest case of uniform grid spacing, W is the identity matrix, with elements $W_{ij} = \delta_{ij}$ where $\delta_{ij} = 1$ only if i=j and is 0 otherwise. For non-uniform grid spacing, it is common to modify this by assuming a diagonal weighting matrix with $W_{ij} = a_i \delta_{ij}$ where a_i is the area associated with the *i'th* grid point. This corresponds to the equal-area weighting discussed by North et al. (1982).

Defining W as a $(p \times p)$ matrix allows for the inclusion of off-diagonal elements. W has the same dimensions as the covariance matrix, and so allows for the weighting to vary not just by grid position, but to vary across the covariance matrix. In some cases more accurate estimates of the area integral can be obtained by using sophisticated numerical integration together with careful spatial smoothing that takes account of the spatial correlations in x^2 (Kagan et al., 1997). The result of such approaches can still be written in the form $x^T W x$, where the effect of smoothing and spatial interpolation is to induce off-diagonal elements in W. The effects of spatially varying errors in x^2 can also be taken into account by including the inverse of an error covariance matrix in W, as is done in optimal interpolation and kriging (Kagan et al., 1997).

It is important to note that any metric W can always be transformed away to Euclidean (uniform weighting) by pre-transforming the grid point variables to new grid point variables

$$x \to x' = W^{1/2} x \tag{3.3}$$

For diagonal metrics, such as the area-weighting metric, $W^{1/2}$ is simply the square root of the diagonal elements of W. For non-diagonal metrics, $W^{1/2}$ can be calculated by using numerical methods such as the Cholesky decomposition or the Denman-Beavers iterative method (Higham, 2008).

The generalized norm can then be written as a Euclidean norm in the new variables:

$$x^T W x = (W^{1/2}x)^T (W^{1/2}x) = x'^T x' \tag{3.4}$$

This is the reason for pre-multiplying anomalies by the *square root* of the grid-box area, as suggested by North et al. (1982). Alternatively, one can find transformations of grid point variables that can remove the metric, for example, by interpolating x to a grid with equal areas (or volumes).

Although the metric W was introduced in order to compensate for unequal grid spacing, the concept can be used in a variety of ways to weight or filter data in EOF analysis. One example is to prewhiten the data in signal-to-noise maximizing EOF analysis (Allen and Smith, 1997; Venzke et al., 1999; Chang et al., 2000). In this case the data matrix is pre-multiplied by a filter (in place of $W^{1/2}$) that is completely determined by the noise covariance matrix. If, for example, the analysis was on a non-uniform grid, the filter matrix and $W^{1/2}$ could be combined by multiplying the data matrix by both, and then performing the EOF analysis.

Another use of the weighting metric is to emphasize (or de-emphasize) regions of the data matrix. This may be desirable due to signal-to-noise ratio, to produce uniform variance across the grid, or simply to eliminate (mask) a region from the analysis. In all these cases, once W is defined, it is used in the same way in the EOF analysis.

3b. Choice of metric to account for sub-grid scale variation

A continuous squared anomaly field, x^2, can be approximated on grid points, but doing so acts as a spatial smoother. This reduction in variance may be small if the decorrelation distance is large, but the reduction can be significant if the field has large small-scale variance. The size of the grid spacing therefore affects the total variance as well as the EOFs. This suggests that EOFs should be calculated on the highest grid resolution available, if one wishes to include the effects of small-scale variance. There may be instances in which it may be desirable to calculate EOFs of large-scale variability [in analogy with EOFs of "low-frequency variability," as in Thompson and Wallace (1998)]. For many data sets the decorrelation distance may be highly variable over different regions of the grid. If EOFs are calculated on a variable grid (e.g., a latitude-longitude grid) that there are two reasons that weighting may be necessary: firstly to compensate for the grid box area, and secondly to compensate for unrepresented variance at a scales smaller than the grid.

Compensation for unrepresented variance can be subtle, in that the coarseness of the grid affects how much of the variance of the continuous field can be represented by the grid. Some of the variance of the continuous field is not represented by the grid point values. In effect, there is variation of the continuous field within each grid box, so that the squares of the data values are underrepresented if the data vary within a grid box.

A continuous anomaly field, x, could vary within a grid box due to the presence of spatial trends across grid points and sub-grid scale random variations in x^2 within grid boxes. However, we can only estimate the area integral of x^2 using grid point values x_i that are considered to be area-averages of the field across the grid box. By Reynolds averaging (Monin and Yaglom, 1971), it can be seen that the area integral of x^2 across a grid box exceeds the area-weighting commonly used in climate studies. If we integrate x^2 over a single grid box,

$$\int_A x^2 da = x_i^2 a_i + \int_A (x - x_i)^2 da \tag{3.12}$$

The first term on the right of (3.12) represents the area weighting discussed in Section 3a. The second term represents variance at scales smaller than the grid box. If the grid is fine enough, so that the gridded values account for nearly all the variance, then the second term is small and area weighting, a_i, would be a good approximation. If the grid is too coarse, then the first term does not adequately account for the total variance of the field.

To illustrate this effect, we consider a single Northern Hemisphere grid of ERA-40 geopotential (Uppala et al., 2005) at 1000 hPa for 1 January 2001. The original resolution is a 1.125° latitude-longitude grid. Figure 1 examines the ratio of $\sum_i x_i^2 a_i$ at lower grid resolutions to $\sum_i x_i^2 a_i$ using the original 1.125° grid. We use two methods for calculating reduced-resolution grid values: 1) interpolation and 2) spatial averages over each grid box. Interpolated values (solid black curve) in Figure 1 illustrate that as the grid resolution is decreased, the mean grid variance is decreased to ~95% at 5-6º and ~90% at 10º. Grid box averaging (gray curve) tends to reduce local minima and maxima in the grid (compared to interpolation) so the falloff is somewhat steeper. Figure 1 illustrates that ~10-20% of the ERA-40 1000-hPa geopotential variance is at grid resolutions smaller than 10º.

In principle, the most extreme case of small-scale variance would be a random anomaly field with variance σ^2 and no spatial correlation between grid points. The area integral would be equal to $\sigma^2 a_i$ but would have a squared grid-box average that scales as $x_i^2 = \sigma^2 / a_i$ (the variance of a mean). In this special case, it would be possible to compensate for the sub-gridscale variance by increasing the weighting factor to a_i^2 rather than a_i (Folland, 1988). Thus, the correct weighting factor for any field would lie between a_i and a_i^2, but for most data fields (e.g., the geopotential used in Figure 1), the factor would be close to a_i.

To further illustrate this effect, consider a uniformly gridded data matrix in which the x_i are uncorrelated with each other and have unit variance. If half of the grid is resampled with grid boxes twice as large, each new grid point value would represent the average of two original values (a_i=2). The variance of the average of two such uncorrelated times series is reduced from 1 to 0.5. Thus, a weighting factor of 2 would be necessary to compensate for the grid box area, and another factor of 2 would be needed to compensate for the reduced variance in the resampled half of the grid, giving a_i^2=4 as the correct weighting, as found by Folland (1988).

Defining a weighting matrix that compensates for unrepresented variance within grid boxes is possible, at least in principle, and would depend on the decorrelation scale and power spectrum of the data at length scales smaller than the grid boxes. For grid spacing smaller than the decorrelation scale of the field (e.g. sea-level pressure on a grid much smaller than the synoptic scale), the correct metric would be $W_{ij} = a_i \delta_{ij}$ (area weighting), whereas for grid boxes containing less coherent variations then the limit of $W_{ij} = a_i^2 \delta_{ij}$ (weighting by the square of the grid box area) could be approached. The correct choice of metric would ensure that the results of the EOF analysis would not be overly dependent on the specific choice of grid.

3c. Generalized EOF analysis

The time-average of the numerical approximation to the area-integral can be written in terms of the sample covariance as follows:

$$\overline{\iint x^2 da} = \iint \overline{x^2} da \approx \frac{1}{n} \sum_{t=1}^{n} x_t^T W x_t = Tr(WS) \qquad (3.11)$$

which defines the generalized total variance. The EOF problem becomes one of finding spatial patterns that maximize $e_i^T W S e_i$ subject to the constraint $e_i^T W e_j = 1$ for $i = j$ and 0 otherwise. By transforming to $e'_i = W^{1/2} e_i$, it can be seen that the EOF solutions are the eigenvectors of the generalized eigenvector equation $SWe_i = \lambda_i e_i$.

Several different approaches can be used to find the generalized EOFs:

1. Solve the generalized eigenvector equation ($SWe_i = \lambda_i e_i$) directly (e.g. using generalized solvers such as the eig function in MATLAB);

2. Transform out the metric by pre-transforming the original variables, find unweighted EOFs, and then back transform the resulting EOFs. Pre-transforming maps the data matrix to $X' = XW^{1/2}$ and the covariance matrix to $S' = (W^{1/2})^T SW^{1/2}$. It is necessary to perform the inverse transformation $e = W^{-1/2}e'$ to obtain the generalized EOFs from the EOFs of the transformed variables;
3. Use iterative projection methods that include W explicitly (Section 4).

4. Projection Methods for Finding EOFs

4a. Projections and EOFs

For any *n*-vector time series of centered anomalies, y, an associated *p*-vector spatial pattern, e, can be obtained by "projecting the data onto the time series," as follows:

$$e = \frac{X^T y}{y^T y} \tag{4.1}$$

where $y^T y = (n-1)s_y^2$ and s_y is the sample standard deviation of the time series. For example, X could be monthly-mean Northern Hemisphere sea level pressure data, and y could be an index of the North Atlantic Oscillation (NAO). (4.1) gives the spatial pattern of the NAO. Note that when $y = x_i$, the time series at grid point *i*, then $e_j = s_{ji}/s_{ii}$ is the covariance map at grid point *i*.

Similarly, a time series y can be obtained from a spatial pattern e by "projecting the data onto the spatial pattern."

$$y = \frac{Xe}{e^T e} \tag{4.2}$$

The metric W does not affect the projection equation 4.1 for obtaining a spatial pattern from a time series, but equation 4.2 for obtaining an EOF time series from an EOF requires weighting of both X and e by $W^{1/2}$.

$$y = \frac{XWe}{e^T We} \tag{4.3}$$

For example, a daily NAO index can be obtained from the NAO spatial pattern and daily sea-level pressure anomalies (e.g., Baldwin et al., 2003) using (4.3). Compensating for a latitude-longitude grid would require the diagonal elements of W to be $\cos\theta$. Note that if the data matrix is pre-transformed by multiplying by $W^{1/2}$, the resulting EOFs will also be transformed. To obtain the EOFs for the original data matrix, either use (4.1) or divide by $W^{1/2}$.

Normalization between the spatial patterns and time series is arbitrary, as long as the product ye^T is preserved. If the spatial patterns are treated as dimensionless weights, then $\frac{1}{Tr(W)} e_1^T W e_1 = 1$. If the EOF time series have unit variance, then ($\frac{1}{n-1} y^T y = 1$). If the EOFs are normalized to have unit normalization, then $\frac{1}{(p-1)Tr(W)} e_1^T W e_1 = 1$.

4b. A simple iterative projection method for calculating EOFs

Several numerical matrix algebra techniques, such as eigenvector analysis and singular value decomposition (SVD) can be used to obtain all the EOFs simultaneously (Jolliffe, 2002). For large data grids, these methods can rapidly become computationally impractical. Although the space and time dimensions in the data matrix can be swapped (von Storch And Zwiers, 1999) even calculating the covariance matrix can be problematic. One alternative is to use SVD directly on the data matrix, rather than the covariance matrix, but the SVD method is still computationally expensive for large data matrices.

Rather than calculating all the EOFs, the first few EOFs and EOF time series can be calculated using iterative techniques (e.g., Holmström, 1963; Clint and Jennings,1970; Jongman et al., 1995; Legendre and Legendre, 1998). One such technique, called the "power method" (Jolliffe, 2002), begins with an initial estimate of the leading EOF, with the only requirement being that the initial estimate has at least some projection onto the leading EOF. The estimated EOF is repeatedly multiplied by the covariance matrix $X^T X$ and normalized during each iteration.

$$e_{k+1} = \frac{X^T X e_k}{\left\| X^T X e_k \right\|} \tag{4.4}$$

The successive estimates converge to the leading EOF as long as the initial estimate has at least minimal projection onto the leading EOF (Golub and van Loan, 1983, p209). The rate of convergence depends on the ratio of the two leading eigenvectors. Thus, an arbitrary initial guess at the EOF is sufficient in practice to initialize the procedure. The EOF time series is never used explicitly, but it would be obtained using (4.2). Once convergence is achieved, the projection of the leading EOF (ye^T) is subtracted from X, and the procedure is repeated to find the desired number of EOFs.

Here we describe another iterative method of finding the leading EOF that involves successive application of (4.1) and (4.3). We will show that this method is an extension of the power method. The main advantages are that this iterative procedure does not require the calculation of the covariance matrix, XX^T, and it accounts for the weighting matrix, W. This method performs the multiplication by XX^T in two steps, and so retains the convergence properties of the power method. The method consists of a simple iterative procedure using the projections (4.1) and (4.3), which alternate between the EOF and its time series. By successively multiplying by X and X^T, we calculate both the EOF and its time series (which was implicit in the power method). Another method involving iteration between a spatial pattern and a time series was proposed by Clint and Jennings (1970). Van den Dool et al. (2000) used a similar approach, without weighting, to find the leading EOF beginning from the leading Empirical Orthogonal Teleconnection (EOT) pattern. Iteration between a time series and spatial pattern to calculate the leading EOF was discovered independently by G. Hegerl (personal communication, 2008).

The method begins with an initial random guess time series whose only requirement is that (as with the power method) it must have a nonzero projection onto the leading EOF time series. Let y_k denote the k-th iteration of the first EOF time series. The data matrix is then projected onto this time series to obtain a spatial pattern, followed by projecting the data matrix onto this spatial pattern to get a new time series. The process is then iterated until the squared error between successive time series falls below a preset tolerance, which can be small but must be greater than the machine precision.

$$e_{k+1} = \frac{X^T y_k}{(y_k)^T y_k} \quad (4.5)$$

$$y_{k+1} = \frac{XWe_{k+1}}{e_{k+1}^T We_{k+1}} \quad (4.6)$$

Iteration continues until convergence, defined by $\|y_{k+1} - y_k\|^2 < \varepsilon$, where ε is a preset tolerance. The denominators in (4.5) and (4.6) are normalization factors. Substituting (4.6) into (4.5) demonstrates that this method is equivalent to the power method. Once convergence is achieved, the EOF multiplied by its time series is subtracted from the data matrix, $X = X - ye^T$, as with the power method, and the process is repeated to find the next EOF. There is no advantage in transposing the data matrix (switching time and space). After convergence the EOF and its time series can be renormalized in any way that preserves the product ye^T.

The algorithm finds one EOF at a time, so it is appropriate to use when only a small number of EOFs are desired. If the eigenvalues of the leading modes are not well separated then convergence will be slow (Golub and van Loan, 1983, P209). Since this iterative approach does not explicitly compute the covariance matrix, it can therefore be used when *X* is a very large matrix. The algorithm does not require significant memory beyond that needed for the data matrix, and the algorithm would work even if the data matrix exceeds computer memory. For example, we have used as a data matrix of 50 years of daily Northern Hemisphere ERA-40 geopotential (dimensions 18,262 by 25,920). The calculation of the leading EOFs was routine using the iterative algorithm.

5. Example

The purpose of this section is to illustrate how to create a weighting matrix, and how different choices for a weighting matrix affect the resulting EOF. It should be emphasized that resulting EOFs may or may not represent physical modes (e.g., Brunet and Vautard, 1996; Ambaum et al., 2001; Jolliffe, 2003). The strict orthogonality condition on EOFs is at the same time its strength and its weakness. For example, the North Atlantic Oscillation may be a more physical mode than the similar EOF of the Northern Annular Mode (Ambaum et al., 2001). We consider the leading EOF of Northern Hemisphere monthly mean zonal mean zonal wind from 1000 to 54 hPa. The data are ECMWF ERA-40 reanalysis (Uppala et al., 2005) from 1958 to 2001 on a 1.125° grid in latitude and unequally spaced pressure levels. The dimensions are n=528 months and with 81 latitudes and 15 pressure levels, p=1215. To remove the annual cycle and the long-term trend, we subtracted the average value for each calendar month and then regressed out the long-term linear trend at each grid point. Figure 2 illustrates the variance $\overline{x^2}$ of the resulting anomalies. The variance is largest in the stratosphere, both near the equator, where the quasi-biennial oscillation dominates (Baldwin et al., 2001), and at high latitudes where month-to-month wintertime changes in zonal wind are large.

This small data set allows us to illustrate how to create three different weighting matrices that varies in both latitude and pressure. We choose W to be a diagonal matrix (one weighting value per grid point), so the diagonal elements of W map to the grid points in the latitude-height plane. To weight in both latitude and pressure, we specify both a latitudinal weighting function, and a vertical weighting function. W is then formed by multiplying these functions. Latitudinal weighting of $\cos\theta$ compensates for the converging meridians at higher latitudes. In the vertical, we choose two different weightings. The first is proportional to the difference in pressure (or mass) between the bottom and top of the layer represented by each level. In the first case, $W_{ij} = \cos\theta_i \Delta p_i \delta_{ij}$. The second case is proportional to the difference in log-pressure (approximately physical height): $W_{ij} = (\cos\theta_i \Delta p_i \delta_{ij}) / p_i$. These weightings are illustrated in Figure 3a and 3b. As a third example, we weight the zonal mean wind to form a quantity proportional to axial relative angular momentum, by pressure weighting in the vertical and $\cos^2\theta$ weighting in latitude (Figure 3c). The global axial relative angular momentum is given by (von Storch, 2000)

$$M_r = \frac{\pi r^3}{4} \int_{-\pi/2}^{\pi/2} \int_0^{P_s} (\overline{u}\cos^2\theta)\, dp\, d\theta \tag{5.1}$$

where r is Earth's radius and P_s is surface pressure. Relative angular momentum is therefore proportional to $\overline{u}\cos^2\theta$ and $W_{ij} = \cos^2\theta_i \Delta p_i \delta_{ij}$. This weighting is illustrated in Figure 3c.

The magnitude of W does not affect the EOFs. W can be multiplied by an arbitrary scaling value without affecting the EOFs. The leading EOFs, corresponding to the weightings in Figure 3, are shown in Figure 4. The EOFs were obtained by iterative projections and include W explicitly in the projection used to find the time series (equations 4.5 and 4.6). The same result could have been obtained by premultiplying by $W^{1/2}$. However, premultiplication by $W^{1/2}$ would be problematic if any of the grid point weights were zero, since the result would then be dividing by $W^{1/2}$. The use of equation 4.5 is more robust, and can be used if some of the grid point weights are zero. The log-pressure weighted EOF (Figure 4a) captures the QBO, while the pressure weighted EOF is dominated by a north-south dipole. The angular momentum EOF (Figure 4c) captures both features. Although the domain shown in Figure 4 is that used for the EOF calculation, equation 4.1 could be used to extend the plots in height or latitude, or to project a different data set onto the EOF time series.

5. Summary

We have developed a generalized technique for including spatial weighting in EOF-type analyses. EOFs, calculated on a set of (possibly irregular) spatial grid points, should not depend on the grid resolution of spacing. To achieve this goal, we introduce a generalized spatial weighting metric into EOF analysis and show how to then find the resulting generalized EOFs. If p is the spatial dimension of the data matrix, then the weighting matrix, W, is dimensioned $p \times p$. Its application is simple and straightforward, involving either 1) pretransforming the data by multiplying by $W^{1/2}$, or carrying W through the EOF calculation.

Although the weighting matrix is developed in order to compensate for an irregular grid, its potential application is much broader. There is a variety of other reasons to spatially weight EOF analysis, including the neglect of unrepresented sub-grid scale variance, masking or emphasizing certain regions, compensating for error covariances, etc. In some cases, the weighting matrix can play the role of a filter. For any of these weighting/filtering strategies, the application to EOF analysis is the same.

The weighting techniques are applicable to all variants of EOF analysis such as extended EOFs, and Maximum Covariance Analysis. All these techniques partition the total variance of the data matrix into discrete modes.

The weighting techniques are developed for gridded data that are equally spaced in time, but the same methodology could be applied to the time domain if the observation times are irregular. The results would be nearly identical to interpolating to a grid that is evenly spaced in time.

Fast, memory-efficient methods are required in order to extract EOFs from large data sets. Often, only the first few EOFs are needed, and with a large dataset it can be computationally more efficient to directly calculate the first few EOFs, rather than use a standard method that finds all the EOFs at once. This study describes one such approach based on a simple iteration of successive projections of the data onto time series and spatial maps. The method is initialized with a random first guess of the EOF time series, and consists of alternately projecting the data matrix onto successive estimates of the EOF and its time series. The technique is a variant of the power method, which has well-defined convergence properties.

The iterative technique works directly from the data matrix (not the covariance matrix), and is suitable for very large data matrices because the covariance matrix is never directly calculated. The technique projects the data onto successive estimates of the EOF spatial pattern and EOF time series. A weighting matrix can be carried through the calculation so that the data are not pre-transformed by multiplying by the square root of appropriate weighting matrices. A subroutine in IDL that implements this method is available from the authors or at www/nwra.com/baldwin.

We argue that the original spatial resolution of the data should be retained when performing any analysis of patterns in geophysical data. Computational issues may lead researchers to degrade the resolution of the data (e.g., by retaining only every other grid point) to make the calculation tractable (Wallace et al., 1992; Wallace, 2006). The degraded resolution reduces the total variance of the field. Since the total variance of the field is reduced, the percent of the variance associated with each EOF can change, resulting in patterns that differ from those that would have been calculated from the full data grid. This could also change the order of the EOFs. Degrading the resolution could therefore affect the determination of whether or not empirically determined EOFs are separate (North et al., 1982; Quadrelli et al., 2005).

Acknowledgements. MPB was funded by NSF (under the US CLIVAR program and the Office of Polar Programs), by NASA's SR&T Program for Geospace Sciences, and Oceans, Ice, & Climate Program. We thank I. Bladé, C.S. Bretherton, G. Brunet, E. Hawkins, G. Hegerl, E. Ioannidou, Y. Kuroda, J. Ochoa, H. van den Dool, J.M. Wallace, and T. Woolings for discussions and comments.

References

Allen, M.R., and L.A. Smith, 1997: Optimal Filtering in Singular Spectrum Analysis, *Phys. Lett. A* **234** (6): 419-428.

Ambaum, M.H.P., Hoskins, B.J. and Stephenson, D.B., 2001: Arctic Oscillation or North Atlantic Oscillation? *J. Climate*, **14**, 3495-3507.

Baldwin, M.P., and K.-K. Tung, 1994: Extra-tropical QBO signals in angular momentum and wave forcing. *Geophys. Res. Lett.*, **21**, 2717-2720.

Baldwin, M.P., L.J. Gray, T.J. Dunkerton, K. Hamilton, P.H. Haynes, W.J. Randel, J.R. Holton, M.J. Alexander, I. Hirota, T. Horinouchi, D.B.A. Jones, J.S. Kinnersley, C. Marquardt, K. Sato, and M. Takahashi, 2001: The quasi-biennial oscillation, *Reviews of Geophysics*, **39** (2), 179-229.

Baldwin, M.P., D.B. Stephenson, D.W.J. Thompson, T.J. Dunkerton, A.J. Charlton, A. O'Neill, 2003: Stratospheric Memory and Skill of Extended-Range Weather Forecasts, *Science*, **301**, 636-640.

Bretherton, C.S., 2003: Empirical Orthogonal Functions and Singular Vectors, Encyclopedia of Atmospheric Sciences, J. R. Holton, J. A. Pyle, and J. Curry, eds., 617-620.

Brunet, G., Empirical normal mode analysis of atmospheric data, 1994: *J. Atmos. Sci.*, **51**, 932-952.

Brunet, G., and R. Vautard, 1996: Empirical Normal Modes versus Empirical Orthogonal Functions for Statistical Prediction, *J. Atmos. Sci.*, **53**, 3468-3489.

Chang, P., R. Saravanan, L. Ji, and G.C. Hegerl, 2000: The Effect of Local Sea-Surface Temperature on Atmospheric Circulation over the Tropical Atlantic Sector. *J. Climate*, **13**, 2195-2216.

Cheng, X., and J. M. Wallace, 1993: Cluster analysis of the Northern Hemisphere wintertime 500-hPa height field: Spatial patterns, *J. Atmos. Sci*, **50**, 2674–2696.

Clint, A., and M. Jennings, 1970: The evaluation of eigenvalues and eigevectors of real symmetric matrices by simultaneous iteration, *Computer J.*, **13**: 76-80.

Craw, I., and P. Cameron, 1992: Face recognition by computer, *Proc. Br. Machine Vision Conf.*, 489-507, Berlin, Springer-Verlag.

Folland, C., 1988: The Weighting of Data in an EOF Analysis, MetO13 Discussion Note No. 113, 4 pp. Met Office, Exeter, UK.

Golub, G.H., and Van Loan, C.F., 1983, Matrix Computations: Baltimore, Md., Johns Hopkins University Press, 476 p.

Hannachi, A., I.T. Jolliffe, D. B. Stephenson, and N. Trendafilov, 2006: In search of simple structures in climate: Simplifying EOFs, *Int. J. Climatology*, **26**, (1), 7-28.

Hannachi, A., I.T. Jolliffe, D.B. Stephenson, 2007: Empirical orthogonal functions and related techniques in atmospheric science, a review, *Int. J. Climatology*, **27**, 1119-1152, DOI 10.1002/joc.1499.

Hawkins, E. and R. Sutton, 2007: Variability of the Atlantic thermohaline circulation described by three-dimensional empirical orthogonal functions, *Climate Dynamics* **29**(7-8): 745-762, doi: 10.1007/s00382-007-0263-8.

Hewistson, B.C., and R.G. Crane, 2002: Self-organized maps: Applications to synoptic climatology, *Clim. Res*., **22**,13-26.

Higham, N.J., 2008: Functions of Matrices: Theory and Computation, Cambridge University Press.

Holmström, I., 1963: On a Method for Parametric Representation of the State of the Atmosphere, *Tellus*, **15**, No. 2, 127-1409.

Jolliffe, I.T., 2002: Principal Component Analysis, 2nd Ed., 487pp, Springer, New York.

Jolliffe, I.T., 2003: A Cautionary Note on Artificial Examples of EOFs, *J. Climate*, 16, 1084-1086, DOI: 10.1175/1520-0442(2003)016.

Jongman, R.H.G., C.J.F ter Braak, and O.F.R. van Tongeren, 1995: Data Analysis in Community and Landscape Ecology, 299pp, Cambridge University Press.

Kagan, R.L., L.S. Gandin, T.M. Smith, 1997: Averaging of Meteorological Fields (Atmospheric and Oceanographic Sciences Library) Springer, 296 pp.

Legendre, P. and L. Legendre, 1998: Numerical Ecology, 2nd edition, Elsevier, 853pp.

Monin, A. S., & A. M., Yaglom, 1971: *Statistical fluid mechanics. Vol 1. Mechanics of turbulence,* MIT Press, Cambridge, MA, 769 pp.

North, G.R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.,* **110,** 600–706.

Quadrelli, R., C.S. Bretherton, and J.M. Wallace, 2005: On sampling errors in empirical orthogonal functions, *J. Climate*, **18**, 3704-3710.

Reusch, D.B., R.B. Alley, and B.C. Hewitson, 2007: North Atlantic climate variability from a self-organizing mp perspective, *J. Geopys. Res.*, **112**, DO2104, doi:10.1029/2006JD007460.

Stephenson, D.B. 1997: Correlation of spatial climate/weather maps and the advantages of using the Mahalanobis metric in predictions, *Tellus*, **49A**, 5, pages 513-527.

Strang, G., 2003: Introduction to Linear Algebra, Third Edition, Wellesley Cambridge Pr, 568 pp.

Thompson, D. W. J. and J. M. Wallace, 1998: The Arctic Oscillation signature in geopotential height and temperature fields, *Geophys. Res. Lett.* **25**(9), 1297–1300, doi:10.1029/98GL00950.10.

Thompson, D.W.J., M.P. Baldwin, and S. Solomon, 2005: Stratosphere/Troposphere coupling in the Southern Hemisphere, *J. Atmos. Sci*., **62**, 708-715.

Uppala, S.M. and co-authors, 2005: The ERA-40 Reanalysis, *Quart. J. Royal. Meteorol. Soc.,* **131**(612), 2961-3012.

van den Dool, H.M., S. Saha, and Å. Johansson, 2000: Empirical Orthogonal Teleconnections, *J. Climate*, **13**, 1421-1435.

van den Dool , H., 2007: Empirical Methods in Short-Term Climate Prediction, Oxford University Press ISBN 0-19-920278-8.

Venzke, S., M.R. Allen, R.T. Sutton and D.P. Rowell, 1999: The atmospheric response over the North Atlantic to decadal changes in sea surface temperature. *J. Climate*, **12**, 2562-2584.

von Storch, H. and F.W. Zwiers, 1999: Statistical Analysis in Climate Research. Cambridge University Press, 484 pp, ISBN 0521 450713.

von Storch, J.-S., 2000: Angular Momenta of the Antarctic and the Arctic Oscillations. *J. Climate.*, **13**, 681-685.

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.

Wallace J. M., Y. Zhang, and K.-H. Lau, 1993a: Structure and seasonality of interannual and interdecadal variability of the geopotential height and temperature fields in the Northern Hemisphere, *J. Climate*, **6**, 2063–2082.

Wallace, J. M., R. L. Panetta, and J. Estberg, 1993b: Representation of the equatorial stratospheric quasi-biennial oscillation in EOF phase space, *J. Atmos. Sci.*, **50**, 1751–1762.

Wallace, J.M., 2006: Class notes on SVD and EOF analyses, http://jisao.washington.edu/wallace/.

List of Figure Captions

Figure 1. Fractional change in Northern Hemisphere average variance of 1000-hPa geopotential on 1 January 2001. The full grid resolution is 1.125°. The black curve shows the total variance at coarser grid resolutions obtained by bi-linear interpolation. The gray curve is obtained by averaging over grid boxes.

Figure 2. Variance of monthly-mean, ERA-40 Northern Hemisphere deseasoned, detrended zonal-mean zonal wind, 1958-2001. The contour intervals are 1, 2.5, 5, 10, 20, 40, 80 m^2s^{-2}. The highest level is 54.6 hPa (one of the ERA-40 assimilation model levels). The data levels correspond to the ticks on the right axis.

Figure 3. Weighting functions corresponding to vertical log-pressure weighting (a), and vertical pressure weighting (b). The latitudinal scaling is by $\cos\theta$ in (a) and (b). (c) corresponds to angular momentum weighting, with latitudinal scaling by $\cos^2\theta$. In all panels the values have been scaled by an arbitrary scaling factor for plotting. The ticks on the right side of the diagrams show the data levels.

Figure 4. Leading EOFs zonal-mean zonal wind with the weighting functions in Figure 2. The data matrix was first multiplied by the square root of the weighting functions. The resulting EOFs were then divided by the square root of the weighting function.

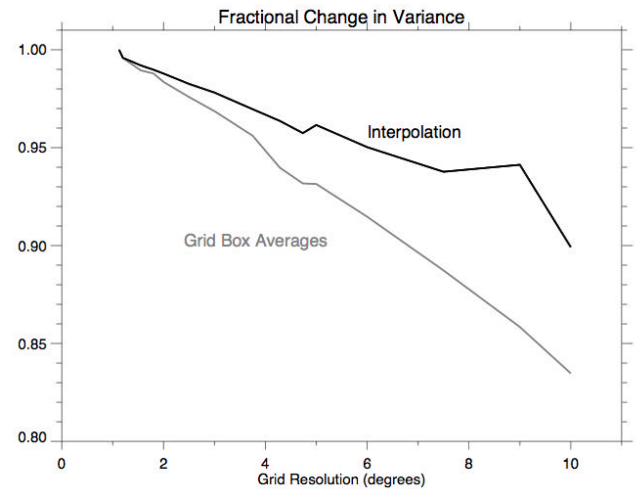

Figure 1. Fractional change in Northern Hemisphere average variance of 1000-hPa geopotential on 1 January 2001. The full grid resolution is 1.125º. The black curve shows the total variance at coarser grid resolutions obtained by bi-linear interpolation. The gray curve is obtained by averaging over grid boxes.

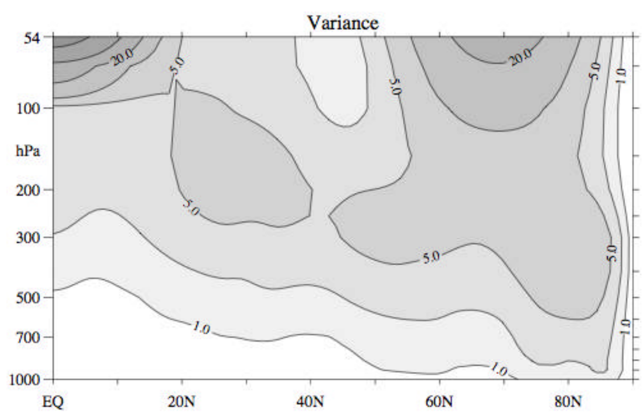

Figure 2. Variance of monthly-mean, ERA-40 Northern Hemisphere deseasoned, detrended zonal-mean zonal wind, 1958-2001. The contour intervals are 1, 2.5, 5, 10, 20, 40, 80 m^2s^{-2}. The highest level is 54.6 hPa (one of the ERA-40 assimilation model levels). The data levels correspond to the ticks on the right axis.

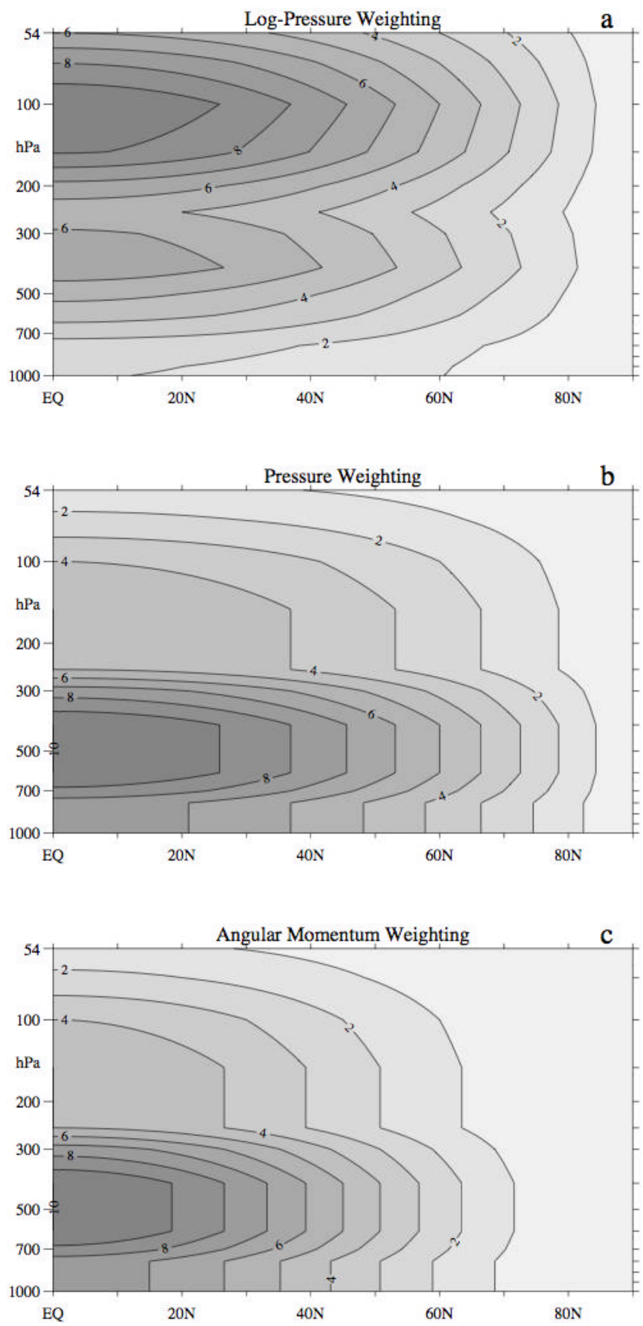

Figure 3. Weighting functions corresponding to vertical log-pressure weighting (a), and vertical pressure weighting (b). The latitudinal scaling is by $\cos\theta$ in (a) and (b). (c) corresponds to angular momentum weighting, with latitudinal scaling by $\cos^2\theta$. In all panels the values have

been scaled by an arbitrary scaling factor for plotting. The ticks on the right side of the diagrams show the data levels.

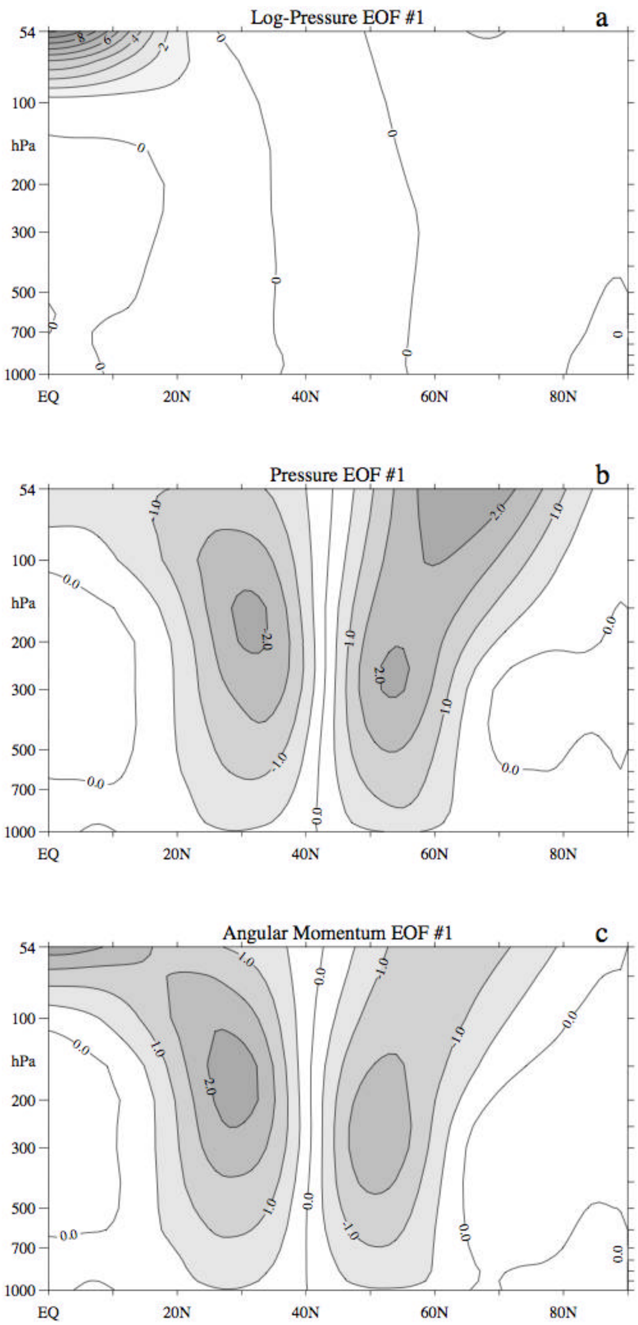

Figure 4. Leading EOFs zonal-mean zonal wind with the weighting functions in Figure 2. The EOFs were obtained using (4.5) and (4.6).