Multivariate Methods in Atmospheric Science

The Sky's the Limit?

Ian Jolliffe

University of Aberdeen

Outline of Talk

- PCA/EOFs - a brief introduction
- Simplifying interpretation
- Extensions to two (or more) groups of variables
- Patterns of behaviour in time (and space)
- Extensions to three or more modes
- Concluding remarks

Simplification

EOFs are often difficult to interpret (though probably less so in atmospheric science than in many other disciplines). To aid in interpretation, various approaches to simplification have been proposed, including the following:

- Orthogonal rotation
- Oblique rotation
- Restriction of loadings to discrete set of values
- Combining variance maximization and simplification criteria
- LASSO-based approach
- Truncation of loadings
- Empirical orthogonal teleconnections

Simplification - Rotation

Well-known and widely used, but controversial (Richman, 1986, 1987; Jolliffe, 1987, 1995; Mestas-Nuñez 2000;)

Among questions to be addressed are:

- orthogonal or oblique
- choice of simplicity criterion
- how many EOFs to rotate
- choice of normalization constraint

Simplification - a discrete set of values for loadings

- Dates back to Hausmann (1982) who allowed only -1, 0, +1

- Vines (2000), Jolliffe *et al.* (2002) simple components allow more integers

- Chipman and Gu (2001) find ordinary EOFs, and then truncate loadings to -1, 0, +1 or to $-c_1, 0, +c_2$

Simplication - combining variance maximization and simplification

Jolliffe and Uddin (2000) successively maximize a criterion which combines variance and simplicity in their Simplified Component Technique (SCoT)

Kiers (1993) goes further and suggests techniques that concentrate on maximizing simplicity, largely ignoring variance maximization.

Simplification - LASSO-based approach

The LASSO (Least Absolute Shrinkage and Selection Operator) approach was developed in multiple regression analysis as a means of dealing with multicollinearity. It is a compromise between variable selection and biased regression. It shrinks some regression coefficients exactly to zero.

Jolliffe and Uddin (2002) adapt the idea to PCA. Adding an extra constraint $\sum_{j=1}^{p} |a_{kj}| < t$ to the usual PCA optimization problem, where a_{kj} is the jth element in the kth EOF and t is a tuning parameter, drives some elements a_{kj} to zero as t decreases.

Simplification - Truncation of Loadings

Setting all loadings whose absolute value is less than some threshold equal to zero is a fairly common informal procedure.

Chipman and Gu (2001) and Richman and Gong (1999) give more formal procedures for doing so.

Truncation should only be done with great caution - Cadima and Jolliffe (1995, 2001) demonstrate that those variables with the smallest loadings are not necessarily the least important in representing a component.

Simplification - Empirical Orthogonal Teleconnections

The patterns in empirical orthogonal teleconnections (van den Dool, 2000) are defined by regressions of the measured variable at each spatial location on its value at a single chosen location. The technique chooses such locations successively, based on how well the location can explain (residual) variation at all other locations.

A complicating factor is that correlations and regressions are calculated on uncentred data.

Relationships between variables in two (or more) groups

Sometimes we wish to relate two or more sets of variables, for example sea surface temperature and mean sea level pressure. Again a variety of techniques is available, including:

- Canonical Correlation Analysis
- Maximum Covariance Analysis
- Redundancy Analysis
- Principal Predictors
- Reduced Rank Regression
- PCA of Instrumental Variables

As we shall see, there are connections between several of these techniques and others.

Two groups of variables - Canonical Correlation Analysis

The oldest amongst such techniques - like PCA was developed by Hotelling in the 1930s - see Gittins (1985).

Given two sets of variables $\mathbf{y}, \mathbf{x}$ of dimensions p_1, p_2, CCA successively finds pairs of linear combinations of the variables $\{\mathbf{a}'_{k1}\mathbf{y}, \mathbf{a}'_{k2}\mathbf{x}\}$ with maximum correlation, subject to being uncorrelated with previous pairs. It turns out that we need to solve the eigen-equation

$$\mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\mathbf{a}_{k2} = l_k\mathbf{R}_{xx}\mathbf{a}_{k2},$$

where $\mathbf{R}_{xy}, \mathbf{R}_{yx}, \mathbf{R}_{xx}, \mathbf{R}_{yy}$ are correlation matrices for appropriate sets of variables.

The solution is invariant to changes in scales of variables (unlike PCA) so the corresponding equation with correlation matrices replaced by covariance matrices leads to an equivalent solution.

Two groups of variables - Maximum Covariance Analysis

Dates back to Tucker (1958), where it was called *inter-battery factor analysis*. Introduced in atmospheric science by Bretherton *et al.* (1992) as singular value decomposition (SVD) – an unfortunate choice of name. The same technique has appeared elsewhere, developed from a number of viewpoints.

One interpretation leads to the better name *maximum covariance analysis* - von Storch and Zweirs (1999, Section 14.1.7). It has also been referred to *canonical covariance analysis* (Cherry, 1997). It is similar in what it does to canonical correlation analysis but has two differences:

- it successively maximizes covariance rather than correlation

- its vectors of loadings are orthogonal, rather than its derived variables being uncorrelated

The optimization problem is solved by finding the SVD of $\mathbf{S}_{yx}$, or equivalently from an eigenanalysis of $\mathbf{S}'_{yx}\mathbf{S}_{yx} = \mathbf{S}_{xy}\mathbf{S}_{yx}$, where $\mathbf{S}_{yx}, \mathbf{S}_{xy}$ are matrices of covariances between $\mathbf{y}$ and $\mathbf{x}$.

Two groups of variables - Redundancy Analysis

In canonical correlation/covariance analyses, the two sets of variables are treated on an equal footing. In redundancy analysis $\mathbf{y}$, $\mathbf{x}$ consist of response and predictor variables respectively. We try to successively maximize the proportion of variance in the response variables that can be accounted for by orthogonal linear combination of the predictor variables (van den Wollenberg, 1977). This leads to the eigen-equation

$$\mathbf{R}_{xy}\mathbf{R}_{yx}\mathbf{a}_{k2} = l_k\mathbf{R}_{xx}\mathbf{a}_{k2}.$$

To find $\mathbf{a}_{k1}$, reverse the rôles of $\mathbf{x}$ and $\mathbf{y}$ in this equation.

Redundancy analysis is equivalent to *PCA of instrumental variables* (Rao, 1964), and to a version of *reduced-rank regression* (Davies and Tso, 1982), as well as having links with multivariate regression.

Two groups of variables - Principal Predictors

Thacker (1999) proposes *principal predictors*. The technique is defined similarly to redundancy analysis, except that the derived variables are uncorrelated rather than the vectors of loadings being orthogonal. This leads to solving the eigen-equation

$$\mathbf{S}_{xy}[diag(\mathbf{S}_{yy})]^{-1}\mathbf{S}_{yx}\mathbf{a}_{k2} = l_k\mathbf{S}_{xx}\mathbf{a}_{k2}.$$

Two groups of variables - other possibilities

- Multivariate Regression, including reduced rank forms - Davies and Tso (1982)

- Regression based on PCs in each group - Preisendorfer and Mobley (1988, Chapter 9)

- CCA based on PCs in each group - Muller (1982)

- Combined PCA of all $p_1 + p_2$ variables - Bretherton *et al.* (1992)

- Partial Least Squares - Frank and Friedman (1993)

- Softly Shrunk Reduced Rank Regression - Aldrin (2000)

- Latent Variable Multivariate Regression - Burnham *et al.* (1999)

- Extensions to more than two groups - see later

Patterns in Time (and Space)

- Climate Change - Fingerprint Techniques
- Detecting Oscillations

Climate Change - Fingerprint Techniques

Given a climate model, parameters within it can be varied and the predicted effects observed. Such parameters can be levels of greenhouse gases, solar output or volcanic activity.

The spatial patterns in predicted climate changes for given parameter changes are sometimes known as fingerprints. When attempting to detect climate change it is often more productive to home in on these fingerprints, and this boils down to trying to maximise a signal/noise ratio.

The problems are high-dimensional, and PCs of the noise covariance matrix may be used instead of the original spatial variables. In one study, North and Wu (2001) suggest keeping 500 PCs out of 3600 - see also Zwiers (1999).

Detecting Oscillations - Singular Spectrum Analysis(SSA)

Also known as singular systems analysis and Pisarenko's method - has books on the subject (Elsner and Tsonis, 1996; Golyandina *et al.*, 2001).

Only one time series, but $(m-1)$ lagged versions of the series are created, and a PCA done on the m variables comprising the series and its lagged versions.

Will detect oscillatory behaviour in the series but tends to find oscillations even when none are present. This can be allowed for (Allen and Smith, 1996).

Detecting Oscillations - Multichannel SSA (MSSA)

Also known as Extended EOF (EEOF) analysis. A total of mp variables are available - the m series from SSA but for p different variables. A PCA is done on all mp variables (Vautard, 1995).

Again oscillatory behaviour may be detected. The p variables may themselves be the first p PCs from a much larger set of variables.

Detecting Oscillations - Principal Oscillation Pattern (POP) Analysis

Given p time series, assume that they follow a first order autoregessive process

$$(\mathbf{x}_{(t+1)}-\mu) = \Upsilon(\mathbf{x}_t-\mu)+\epsilon_t,\ t = 1, 2, \ldots, (n-1).$$

POP analysis uses multivariate regression analysis to estimate Υ, and then finds eigenvectors of this estimated matrix of regression coefficients. The eigenvectors are known as *principal oscillation patterns*. Some of these eigenvectors are complex and represent damped oscillations (von Storch *et al.*, 1988).

As with other techniques, the data may first be transformed to PCs before POP analysis is undertaken.

Detecting Oscillations - Hilbert (Complex) EOFs

The real p-variate time series $\mathbf{x}_t$ is made complex by appending as its imaginary part the Hilbert transform of the series, and a PCA is carried out on this complex series.

The definition of Hilbert transform and how to estimate it for finite series need not concern us - see von Storch and Zwiers (1999, Section 16.2.4). However, if $\mathbf{x}_t$ is made up of oscillatory terms, the Hilbert transform advances each term by $\pi/2$ radians.

There is a connection between HEOF analysis and PCA in the frequency domain.

Detecting Oscillations - other possibilities

- Kooperberg and O'Sullivan (1996) describe a hybrid of PCA and POP analysis, leading to what they call *Predictive Oscillation Patterns* (PROPs)

- Multitaper frequency-domain singular value decomposition (Mann and Park, 1999)

- Cyclo-stationary and Periodically Extended EOFs (and POPs) - Kim and Wu (1999)

Extensions to three (or more) modes

A typical atmospheric science application has the two 'modes' of time and space. There may be a third mode, corresponding to different variables measured at each time-space location, or to different levels of the atmosphere. We may even have all four modes available. There are a number of extensions of PCA to deal with three-mode or multi-way data, including:

- O-mode, P-mode, ... , T-mode analysis
- Extended EOF analysis
- Three mode PCA

Three modes - O-mode ... T-mode analysis

Not really an extension. If there are 3 modes (space, time, measurement), we can choose 2 out of 3 to represent the variables and observations respectively in a PCA, and analyse those two fixing the third, in 6 ways. O-mode, P-mode, Q-mode, R-mode, S-mode, T-mode correspond to these 6 possibilities (Cattell, 1978; Richman, 1986). S-mode (locations = variables, times = observations) is most common in atmospheric science, but T-mode (times = variables, locations = observations) and occasionally others are also sometimes used.

Three modes - Extended EOF analysis

If there are n times, s locations, and p variables measured at each space-time location, we combine locations and variables to give a total of sp variables in a $(n \times sp)$ data matrix, and perform PCA on that matrix. It would be possible to combine two of the three modes in two other ways. It is also possible to incorporate data at different time lags, leading multivariate Extended EOF analysis (Mote *et al.*, 2000). The latter also extends multichannel singular spectrum analysis (MSSA - Plaut and Vautard, 1994).

Some Concluding Remarks

- In all the techniques, especially those concerned with simplification, there is a desire to find physical interpretations for the reults. This is frequently controversial - often there is a failure to recognise that the criteria underlying the techniques are not designed to look for physically meaningful results. If the physical 'modes' are already known, the analysis may be redundant.

- An example of this is the conflicting arguments that 'physical modes' are likely to be correlated (oblique rotation), or that they should be independent (independent component analysis - Aires *et al.* (2000)).