Chapter 13

Spatial Patterns: EOFs and CCA

by Hans von Storch

13.1 Introduction

Many analyses of climate data sets suffer from high dimensions of the variables representing the state of the system at any given time. Often it is advisable to split the full phase space into two subspaces. The "signal" space is spanned by few characteristic patterns and is supposed to represent the dynamics of the considered process. The "noise subspace", on the other hand, is high-dimensional and contains all processes which are purportedly irrelevant in their details for the "signal subspace".

The decision of what to call"signal" and what to call "noise" is non-trivial. The term "signal" is not a well-defined expression in this context. In experimental physics, the signal is well defined, and the noise is mostly the uncertainty of the measurement and represents merely a nuisance. In climate research, the signal is defined by the interest of the researcher and the noise is everything else unrelated to this object of interest. Only in infrequent cases is the noise due to uncertainties of the measurement, sometimes the noise comprises the errors introduced by deriving "analyses", i.e., by deriving from many irregularly distributed point observations a complete map. But in most cases the noise is made up of well-organized processes whose

Acknowledgments: I am grateful to Victor Ocaña, Gabriele Hegerl, Bob Livezey and Robert Vautard for their most useful comments which led to a significant (not statistically meant) improvement of the manuscript. Gerassimos Korres supplied me with Figures 13.1 and 13.2.

details are unimportant for the "signal". In many cases the noise is not a nuisance but its statistics are relevant for the understanding of the dynamics of the signal (see also Chapter 3). Generally the signal has longer scales in time and space than the noise, and the signal has fewer degrees of freedom than the noise.

An example is oceanic heat transport - this signal is low frequent and large in spatial scale. The extratropical storms are in this context noise, since the individual storms do not matter, but the ensemble of the storms, or the *storm track* is of utmost importance as this ensemble controls the energy exchange at the interface of atmosphere and ocean. Thus, for some oceanographers the individual storms are noise. For a synoptic meteorologist an individual storm is the object of interest, and thus the signal. But to understand an individual storm does not require the detailed knowledge of each cloud within the storm, so the clouds are noise in this context.

The purpose of this chapter is to discuss how the "signal" subspace may be represented by characteristic patterns. The specification of such characteristic patterns can be done in various ways, ranging from purely subjectively defined patterns, patterns with favorable geometric properties like a powerful representation of prescribed spatial scales (such as spherical harmonics) to patterns which are defined to optimize statistical parameters. Empirical Orthogonal Functions (EOFs) are optimal in representing variance; Canonical Correlation Patterns (CCPs) maximize the correlation between two simultaneously observed fields); others such as PIPs and POPs (see Chapter 15) satisfy certain dynamical constraints.

In this contribution we first represent the general idea of projecting large fields on a few "guess patterns" (Section 13.2). Then EOFs are defined as those patterns which are most powerful in explaining variance of a *random field* $\vec{\mathbf{X}}$ (Section 13.3). In Section 13.4 the *Canonical Correlation Analysis* of two simultaneously observed random fields $(\vec{\mathbf{X}}, \vec{\mathbf{Y}})$ is introduced. In Section 13.5 two methods of determining patterns optimized to represent maxima of variance are sketched, namely *Empirical Orthogonal Teleconnections* and *Redundancy Analysis*.

13.2 Expansion into a Few Guess Patterns

13.2.1 Guess Patterns, Expansion Coefficients and Explained Variance

The aforementioned separation of the full phase space into a "signal" subspace, spanned by a few patterns $\vec{p}^{\,k}$ and a "noise" subspace may be formally written as

$$\vec{\mathbf{X}}_t = \sum_{k=1}^{K} \alpha_k(t)\vec{p}^{\,k} + \vec{n}_t \tag{13.1}$$

with t representing in most cases time. The K "guess patterns" $\vec{p}^{\,k}$ and time coefficients $\alpha_k(t)$ are supposed to describe the dynamics in the signal subspace, and the vector $\vec{n}_t$ represents the "noise subspace". When dealing with the expressions "signal" and "noise" one has to keep in mind that "noise subspace" is implicitly defined as that space which does not contain the "signal". Also, since the noise prevails everywhere in the phase space, a complete separation between "signal" and "noise" is impossible. Indeed, the "signal subspace" contains an often considerable amount of noise.

The truncated vector of state $\vec{\mathbf{X}}_t^{\,S} = \sum_{k=1}^{K} \alpha_k(t)\vec{p}^{\,k}$ is the projection of the full vector of state on the signal subspace. The residual vector $\vec{n}_t = \vec{\mathbf{X}}_t - \vec{\mathbf{X}}_t^{\,S}$ represents the contribution from the noise subspace.

The vector $\vec{\mathbf{X}}$ is conveniently interpreted as a *random vector* with expectation $\mathrm{E}\left(\vec{\mathbf{X}}\right) = \vec{\mu}$, covariance matrix $\mathbf{\Sigma} = \mathrm{E}\left((\vec{\mathbf{X}} - \vec{\mu})(\vec{\mathbf{X}} - \vec{\mu})^T\right)$ and the variance $\textsc{Var}\left(\vec{\mathbf{X}}\right) = \sum_i \mathrm{E}\big((\mathbf{X}_i - \mu_i)^2\big)$. Then, the expansion coefficients α_k are univariate random variables whereas the patterns are constant vectors. In the case of EOFs, CCA and similar techniques the patterns are derived from $\vec{\mathbf{X}}$ so that they represent *parameters* of the random vector $\vec{\mathbf{X}}$.

The expansion coefficients[1] $\vec{\alpha} = (\alpha_1, \alpha_2 \ldots \alpha_K)^T$ are determined as those numbers which minimize

$$\epsilon(\vec{\alpha}) = \langle \vec{\mathbf{X}} - \textstyle\sum_k \alpha_k \vec{p}^{\,k}, \vec{\mathbf{X}} - \sum_k \alpha_k \vec{p}^{\,k} \rangle \tag{13.2}$$

with the "dot product' $\langle \vec{a}, \vec{b} \rangle = \sum_j a_j b_j$. The optimal vector of expansion coefficients is obtained as a zero of the first derivative of ϵ:

$$\sum_{i=1}^{K} \vec{p}^{\,k\,T} \vec{p}^{\,i} \alpha_i = \vec{p}^{\,k\,T} \vec{\mathbf{X}} \tag{13.3}$$

After introduction of the notation $\mathcal{A} = \left(\vec{a}|\vec{b}\cdots\right)$ for a matrix $\mathcal{A}$ with the first column given by the vector $\vec{a}$ and the second column by a vector $\vec{b}$, (13.3) may be rewritten as

$$\mathcal{P}\vec{\alpha} = \left(\vec{p}^{\,1}|\cdots|\vec{p}^{\,K}\right)^T \vec{\mathbf{X}} \tag{13.4}$$

with the symmetric $K \times K$-matrix $\mathcal{P} = \left(\vec{p}^{\,k\,T}\vec{p}^{\,i}\right)$. In all but pathological cases the matrix $\mathcal{P}$ will be invertible such that a unique solution of (13.3) exists:

$$\vec{\alpha} = \mathcal{P}^{-1} \left(\vec{p}^{\,1}|\cdots|\vec{p}^{\,K}\right)^T \vec{\mathbf{X}} \tag{13.5}$$

Finally, if we define K vectors $\vec{p}_A^{\,1} \ldots \vec{p}_A^{\,K}$ so that

$$\left(\vec{p}_A^{\,1}|\cdots|\vec{p}_A^{\,K}\right) = \left(\vec{p}^{\,1}|\cdots|\vec{p}^{\,K}\right)\mathcal{P}^{-1} \tag{13.6}$$

[1] The expansion coefficients may be seen as the transformed coordinates after introducing the guess patterns as new basis of the phase space.

Then the k-th expansion coefficient α_k is given as the dot product of the vector of state $\vec{\mathbf{X}}$ and the "adjoint pattern" $\vec{p}_A^{\,k}$:

$$\alpha_k = \langle \vec{p}_A^{\,k}, \vec{\mathbf{X}} \rangle \tag{13.7}$$

In some cases, and in particular in case of EOFs, the patterns $\vec{p}^{\,k}$ are *orthogonal* such that $\mathcal{P}$ is the identity matrix and $\vec{p}^{\,k} = \vec{p}_A^{\,k}$. In this case (13.7) reads

$$\alpha_k = \langle \vec{p}^{\,k}, \vec{\mathbf{X}} \rangle \tag{13.8}$$

A convenient measure to quantify the relative importance of one pattern or of a set of patterns $\{\vec{p}^{\,k}\}$ is the "amount of explained variance", or, more precisely, the "proportion of variance accounted for by the $\{\vec{p}^{\,k}\}$" [formally similar to the "Brier-based score" β introduced in (10.6)]:

$$\eta = \frac{\mathrm{VAR}\Big(\vec{\mathbf{X}}\Big) - \mathrm{VAR}\Big(\vec{\mathbf{X}} - \sum_k \alpha_k \vec{p}^{\,k}\Big)}{\mathrm{VAR}\Big(\vec{\mathbf{X}}\Big)} \tag{13.9}$$

where we have assumed that the random vector $\vec{\mathbf{X}}$ would have zero mean ($\mathrm{E}\Big(\vec{\mathbf{X}}\Big) = 0$). If the data are not centered then one may replace the variance-operator in (13.9) by the sum of second moments $\mathrm{E}\Big(\vec{\mathbf{X}}^T \vec{\mathbf{X}}\Big)$ and refer to the explained second moment.

The numerical value of the explained variance is bounded by $-\infty < \eta \leq 1$.[2] If $\eta = 0$ then $\mathrm{VAR}\Big(\vec{\mathbf{X}} - \sum_k \alpha_k \vec{p}^{\,k}\Big) = \mathrm{VAR}\Big(\vec{\mathbf{X}}\Big)$ and the representation of $\vec{\mathbf{X}}$ by the patterns is useless since the same result, in terms of explained variance, would have been obtained by arbitrary patterns and $\alpha_k = 0$. On the other end of the scale we have $\eta = 1$ which implies $\mathrm{VAR}\Big(\vec{\mathbf{X}} - \sum_k \alpha_k \vec{p}^{\,k}\Big) = 0$ and thus a perfect representation of $\vec{\mathbf{X}}$ by the guess patterns $\vec{p}^{\,k}$.

The amount of explained variance, or, sloppily formulated, "the explained variance", can also be defined locally for each component j:

$$\eta(j) = 1 - \frac{\mathrm{VAR}\big(\mathbf{X}_j - \sum_k \alpha_k p_j^k\big)}{\mathrm{VAR}(\mathbf{X}_j)} \tag{13.10}$$

If the considered random vector $\vec{\mathbf{X}}$ can be displayed as a map then also the amount of explained variance η can be visualized as a map.

[2] The number η can indeed be negative, for instance when $\sum_k \alpha_k \vec{p}^{\,k} = -\vec{\mathbf{X}}_t$. Then, $\eta = -3$.

13.2.2 Example: Temperature Distribution in the Mediterranean Sea

As an example[3], we present here an expansion (13.1) of a time-dependent 3-dimensional temperature field of the Mediterranean Sea. The output of a 9-year run of an OGCM, forced by monthly mean atmospheric conditions as analysed by the US-NMC for the years 1980 to 1988, was decomposed such that

$$\vec{\mathbf{T}}(\vec{r}, z, t) = \sum_k \alpha_k(z, t)\vec{p}^k_{\vec{r}} \tag{13.11}$$

with $\vec{r}$ representing the horizontal coordinates, z the vertical coordinate and t the time. The temperature field is given on a $(\vec{r}, z)$-grid with better resolution in the upper levels. In the representation (13.11), coefficients α_k depend on depth and time. The orthogonal patterns $\vec{p}^{\,k}$ depend only on the horizontal distribution and are independent of the depth z and of the time t. The decomposition was determined by a "Singular Value Decomposition" but the technical aspects are not relevant for the present discussion.

Prior to the analysis, the data have been processed. For each depth z the horizontal spatial mean and standard deviation have been calculated. Then, the temperature values at each depth are normalized by subtracting the (spatial) mean and dividing by the (spatial) standard deviation of the respective depth. The annual cycle is *not* subtracted so that the time series are not stationary (but cyclo-stationary) with an annual cycle of the time mean and of the temporal standard deviation.

The first two patterns are shown in Figure 13.1. The first one, which represents 57% of the total second moment of the normalized temperature, exhibits a dipole, with about half of the basin being warmer than average and the other half being cooler than average. The second mode, which represents 32% of the second moment, is relatively uniform throughout the Mediterranean Sea. The relative importance of the two modes for different layers of the ocean is described by the amount of the 2nd moment accounted for by the two modes (Figure 13.2). The second mode explains most of the 2nd moment above 100 m, whereas the first mode dominates below the top 4 layers. An inspection of the time series $\alpha_k(z, t)$ for different depths z reveals that the two modes represent different aspects of the climatology of the Mediterranean Sea. The time series $\alpha_1(z, t)$ is always positive with irregular variations superimposed. Such a behavior is indicative that the first mode describes mostly the overall mean and its interannual variability. The time series $\alpha_2(z, t)$ describe a regular annual cycle (with a negative minimum in winter so that $\alpha_2\vec{p}^1_{\vec{r}}$ is a negative distribution; and a positive maximum in

[3]This material was presented by Gerrasimos Korres in a "student paper" during the Autumn School. It will be available as a regular paper co-authored by Korres and Pinardi in 1994.

summer - indicating warmer than average conditions) plus a slight upward trend (gradual warming).

13.2.3 Specification of Guess Patterns

There are various ways to define the patterns $\vec{p}^{\,k}$:[4]

- The very general approach is Hasselmann's "Principal Interaction Patterns" formulation (PIP; Hasselmann, 1988). The patterns are implicitly defined such that their coefficients $\alpha_k(t)$ approximate certain dynamical equations, which feature unknown parameters.
- A simplified version of the PIPs are the "Principal Oscillation Patterns" (POPs, H. von Storch et al., 1988, 1993), which model linear dynamics and which have been successfully applied for the analysis of various processes (see Chapter 15).
- A standard statistical exercise in climate research aims at the identification of expected signals, such as the atmospheric response to enhanced greenhouse gas concentrations or to anomalous sea-surface temperature conditions (see Chapter 8). This identification is often facilitated by the specification of patterns determined in experiments with general circulation models (H. von Storch, 1987; Santer et al., 1993, Hegerl et al., 1996).

[4] See also Section 8.3.2.

Also patterns "predicted" by simplified dynamical theory (for instance, linear barotropic equations) are in use (Hannoschöck and Frankignoul, 1985; Hense et al., 1990).

- A frequently used class of patterns are orthogonal functions such as trigonometric functions or spherical harmonics. In all "spectral" atmospheric general circulation models the horizontal fields are expanded according to (13.1) with spherical harmonics as guess patterns. In the spectral analysis of time series the trigonometric functions are used to efficiently represent fields.

- The Empirical Orthogonal Functions (EOFs) and Canonical Correlation Patterns (CCPs) are very widely used guess patterns. These choices will be discussed in some length in the next two Sections 13.3 and 13.4. Offsprings of these techniques are *Extended EOFs* (EEOFs) and *Complex EOFs* (CEOFs). In the EEOFs (Weare and Nasstrom, 1982; see also Chapter 14) the same vector at *different* times is concatenated; in the CEOF (Wallace and Dickinson, 1972; Barnett, 1983; also Section 15.3.4) the original vector real-valued time series is made complex by adding its *Hilbert transform* as imaginary component. The Hilbert transform may be seen as a kind of "momentum". Both techniques are successfully applied in climate research but we will not go into details in the present review. They are presented in more detail in H. von Storch and Zwiers (1999). A variant of Canonical Correlation Analysis is Redundancy Analysis and was proposed by Tyler (1982). It identifies pairs of patterns, so that a maximum of variance of the predictand is obtained through regression from the predictor. We will briefly introduce this technique in Section 13.5.2.

- A new approach named Empirical Orthogonal Teleconnections (EOTs), proposed by van den Dool et al. (2000), constructs stepwise orthogonal patterns by determining points with maximum skill in linearly specifying all other data points. This technique is sketched in Section 13.5.1. EOTs may be seen as a variant of conventional teleconnections patterns (see Chapter 12), which allows to split up the total variance into a sum of contributions from a limited number of patterns.

- The "wavelet" analysis is a technique which projects a given time series on a set of patterns, which are controlled by a location and a dispersion parameter. See Meyers et al. (1993) or Farge et al. (1993).

13.2.4 Rotation of Guess Patterns

For EOFs there exists a widely used variant named *Rotated EOFs* (for instance, Barnston and Livezey, 1987). The name is somewhat misleading as it indicates that the "rotation" would exploit properties special to the EOFs.

This is not the case. Instead, the general concept of "rotation" is to replace the patterns $\vec{p}^{\,k}$ in (13.1) by "nicer" patterns $\vec{p}_R^{\,k}$:

$$\sum_{k=1}^{K} \alpha_k \vec{p}^{\,k} = \sum_{k=1}^{K} \alpha_k^R \vec{p}_R^{\,k} \tag{13.12}$$

The patterns $\vec{p}_R^{\,k}$ are determined such that they minimize a certain (nonlinear) functional of "simplicity" F_R and that they span the same space as the original set of vectors $\{\vec{p}^{\,k}\}$. Constraints like unit length ($\vec{p}_R^{\,k\,T}\vec{p}_R^{\,k} = 1$) and, sometimes, orthogonality ($\vec{p}_R^{\,k\,T}\vec{p}_R^{\,i} = 0$) are invoked. Richman (1986) lists five vague criteria for patterns being "simple" and there are many proposals of "simplicity" functionals. If the patterns are not orthogonal the term *oblique* is used. The minimization of functionals such as (13.13) is in general non-trivial since the functionals are nonlinear. Numerical algorithms to approximate the solutions require the number of involved dimensions K to be not too large.

A widely used method is the "varimax", which generates a set of orthogonal patterns which minimize the joint "simplicity" measure

$$F_R(\vec{p}_R^{\,1} \cdots \vec{p}_R^{\,K}) = \sum_{k=1}^{K} f_R(\vec{p}_R^{\,k}) \tag{13.13}$$

with functions f_R such as

$$f_R(\vec{p}) = \frac{1}{m}\sum_{i=1}^{m}\left(p_i^2\right)^2 - \frac{1}{m^2}\left(\sum_{i=1}^{m} p_i^2\right)^2 \qquad \text{or,} \tag{13.14}$$

$$f_R(\vec{p}) = \frac{1}{m}\sum_{i=1}^{m}\left[\left(\frac{p_i}{s_i}\right)^2\right]^2 - \frac{1}{m^2}\left[\sum_{i=1}^{m}\left(\frac{p_i}{s_i}\right)^2\right]^2 \tag{13.15}$$

The number p_i is the ith component of a m-dimensional vector $\vec{p}$, s_i is the standard deviation of the ith component of $\vec{\mathbf{X}}^S$, which is the projection of the original full random vector $\vec{\mathbf{X}}$ in the signal subspace spanned by the K vectors $\{\vec{p}^{\,1} \cdots \vec{p}^{\,K}\}$.

Both definitions (13.14,13.15) have the form of a variance: in the "raw varimax" set-up (13.14) it is the (spatial) variance of the squares of the components of the pattern $\vec{p}$ and in the "normal varimax" (13.15) it is the same variance of a normalized version $\vec{p}\,' = (p_i/s_i)$ with $(s_1^2 \cdots s_m^2)^T = \textsc{Var}\left(\mathbf{X}_i^S\right) = \sum_{k=1}^{K} \alpha_k p_i^k$. Minimizing (13.13) implies therefore finding a set of K patterns $\vec{p}_R^{\,k}$ such that their squared patterns have (absolute or relative) minimum spatial variance. The functions f_R are always positive and are zero if all $p_i = 0$ or 1 (13.14) or if all $p_i = s_i$ (13.15).

The results of a rotation exercise depend on the number K and on the choice of the measure of simplicity. The opinion in the community is divided

on the subject of rotation. Part of the community advocates the use of rotation fervently as a means to define physically meaningful, statistically stable patterns whereas others are less convinced because of the hand-waving character of specifying the simplicity functions, and the implications of this specification for the interpretation of the result. The successful application of the rotation techniques needs some experience and it might be a good idea for the novice to have a look into Richman's (1986) review paper on that topic. Interesting examples are offered by, among many others, Barnston and Livezey (1987) and Chelliah and Arkin (1992). Cheng et al. (1995) found that a conventional EOF analysis yields statistically less stable patterns than a rotated EOF analysis.

In the present volume, Section 6.3.5 is dealing with a varimax-rotation (13.14) of a subset of EOFs.

13.3 Empirical Orthogonal Functions

For the sake of simplicity we assume in this Section that the expectation of the considered random vector $\vec{\mathbf{X}}$ is zero: $\vec{\mu} = 0$. Then the covariance matrix of $\vec{\mathbf{X}}$ is given by $\mathbf{\Sigma} = \mathrm{E}\left(\vec{\mathbf{X}}\vec{\mathbf{X}}^T\right)$.

The vector $\vec{\mathbf{X}}$ may represent very different sets of numbers, such as

- Observations of different parameters at one location (such as daily mean temperature, sunshine, wind speed etc.)
- Grid-point values of a continuous field which was spatially discretized on a regular grid (as is often the case for horizontal distributions) or on an irregular grid (such as the vertical discretization in GCMs).
- Observations of the same parameter (such as temperature) at irregularly distributed stations (see the example of Central European temperature in Section 13.3.4; also Briffa dealt with this case in Section 5.6.1 when he considered tree ring data from different sites).

There is some confusion with the terms, since several alternative sets of expressions are in use for the same object. What is labeled an EOF here is also named a *principal vector* or a *loading*, whereas *EOF coefficients* are sometimes *principal components* (for instance in Chapter 8) or *scores*.[5]

13.3.1 Definition of EOFs

Empirical Orthogonal Functions are defined as that set of K *orthogonal* vectors (i.e., ${\vec{p}^{\,k}}^T \vec{p}^{\,i} = \delta_{ki}$) which minimize the variance of the residual $\vec{n}$ in

[5] The expressions "principal vector" stems from the geometrical interpretation that these vectors are the principal vectors of an ellipsoid described by the covariance matrix. The terms "loading" and "scores" come from factor analysis, a technique widely used in social sciences. The term "EOF" seems to be in use only in meteorology and oceanography.

(13.1).[6] Because of the enforced orthogonality the coefficients $\vec{\alpha}$ are given by (13.8).

The EOFs are constructed consecutively: In a first step the pattern $\vec{p}^{\,1}$ of unit length (${\vec{p}^{\,1}}^T\vec{p}^{\,1} = 1$) is identified which minimizes

$$\mathrm{E}\left(\left(\vec{\mathbf{X}} - \alpha_1\vec{p}^{\,1}\right)^T\left(\vec{\mathbf{X}} - \alpha_1\vec{p}^{\,1}\right)\right) = \epsilon_1 \tag{13.16}$$

After the first EOF $\vec{p}^{\,1}$ is determined, the second EOF is derived as the pattern minimizing

$$\mathrm{E}\left(\left[\left(\vec{\mathbf{X}} - \alpha_1\vec{p}^{\,1}\right) - \alpha_2\vec{p}^{\,2}\right]^T\left[\left(\vec{\mathbf{X}} - \alpha_1\vec{p}^{\,1}\right) - \alpha_2\vec{p}^{\,2}\right]\right) = \epsilon_2 \tag{13.17}$$

with the constraints ${\vec{p}^{\,1}}^T\vec{p}^{\,2} = 0$ and ${\vec{p}^{\,2}}^T\vec{p}^{\,2} = 1$. In similar steps the remaining EOFs are determined. A K-dimensional field has in general K EOFs but we will see below that in practical situations the number of EOFs is limited by the number of samples.

We demonstrate now how to get the first EOF. The derivation of the other EOFs is more complicated but does not offer additional significant insights (for details, see H. von Storch and Hannoschöck, 1986). Because of the orthogonality we may use (13.8) and reformulate (13.16) such that

$$\begin{aligned}\epsilon_1 &= \mathrm{E}\left(\vec{\mathbf{X}}^T\vec{\mathbf{X}}\right) - 2\mathrm{E}\left(\left(\vec{\mathbf{X}}^T\vec{p}^{\,1}\right){\vec{p}^{\,1}}^T\vec{\mathbf{X}}\right) + \mathrm{E}\left(\vec{\mathbf{X}}^T\vec{p}^{\,1}\vec{\mathbf{X}}^T\vec{p}^{\,1}\right)\\ &= \mathrm{VAR}\left(\vec{\mathbf{X}}\right) - {\vec{p}^{\,1}}^T\mathbf{\Sigma}\vec{p}^{\,1}\end{aligned} \tag{13.18}$$

To find the minimum, a Lagrange multiplier λ is added to enforce the constraint ${\vec{p}^{\,1}}^T\vec{p}^{\,1} = 1$. Then the expression is differentiated with respect to $\vec{p}^{\,1}$ and set to zero:

$$\mathbf{\Sigma}\vec{p}^{\,1} - \lambda\vec{p}^{\,1} = 0 \tag{13.19}$$

Thus, the first (and all further) EOF must be an eigenvector of the covariance matrix $\mathbf{\Sigma}$. Insertion of (13.19) into (13.18) gives

$$\epsilon_1 = \mathrm{VAR}\left(\vec{\mathbf{X}}\right) - \lambda \tag{13.20}$$

so that a minimum ϵ_1 is obtained for the eigenvector $\vec{p}^{\,1}$ with the largest eigenvalue λ.

More generally, we may formulate the **Theorem**:

[6] The approach of minimizing the variance of the residual has a mathematical background: the variance of the residual is a measure of the "misfit" of $\vec{\mathbf{X}}$ by $\sum_{k=1}^{K}\alpha_k(t)\vec{p}^{\,k}$. Other such measures of misfit could be chosen but this quadratic form allows for a simple mathematical solution of the minimization problem.

The first K eigenvectors $\vec{p}^{\,k}$, for any $K \leq m$, of the covariance matrix $\mathbf{\Sigma} = \mathrm{E}\left(\vec{\mathbf{X}}\vec{\mathbf{X}}^T\right)$ of the m-variate random vector $\vec{\mathbf{X}}$ form a set of pairwise orthogonal patterns. They minimize the variance

$$\epsilon_K = \mathrm{E}\left(\left(\vec{\mathbf{X}} - \sum_{k=1}^{K} \alpha_k \vec{p}^{\,k}\right)^2\right) = \mathrm{VAR}\left(\vec{\mathbf{X}}\right) - \sum_{k=1}^{K} \lambda_k \tag{13.21}$$

with $\alpha_k = \vec{\mathbf{X}}^T \vec{p}^{\,k}$ and $\vec{p}^{\,k\,T}\vec{p}^{\,k} = 1$. The patterns are named "Empirical Orthogonal Function".

From the construction of the EOFs it becomes clear the patterns represent an optimal potential to compress data into a minimum number of patterns. *Sometimes* the first EOF of the first few EOFs represent a meaningful physical summary of relevant processes, which go with characteristic patterns. For further discussion see Section 13.3.2

Favorable aspects of the EOFs are the geometrical orthogonality of the patterns and the *statistical independence*, or, more correctly, the zero-correlation of the "EOF coefficients" α_i:

$$\mathrm{E}(\alpha_i \alpha_k) = \vec{p}^{\,i\,T}\mathrm{E}\left(\vec{\mathbf{X}}\vec{\mathbf{X}}^T\right)\vec{p}^{\,k} = \vec{p}^{\,i\,T}\mathbf{\Sigma}\vec{p}^{\,k} = \lambda_k \vec{p}^{\,i\,T}\vec{p}^{\,k} = \lambda_k \delta_{k,i} \tag{13.22}$$

A byproduct of the calculation (13.22) is $\mathrm{VAR}(\alpha_k) = \lambda_k$.

EOFs are *parameters* of the random vector $\vec{\mathbf{X}}$. If $\vec{\mathbf{X}}$ is a Gaussian distributed random vector then the set of coefficients α_k form a set of univariate normally distributed independent random variables (with zero means and standard deviations given by the square root of the respective eigenvalues, if the mean of $\vec{\mathbf{X}}$ is zero.)

The relative importance of the EOFs may be measured by their capability to "explain" $\vec{\mathbf{X}}$-variance. This amount of explained variance η can be calculated for individual EOFs or for sets of EOFs, for the complete m-variate vector $\vec{\mathbf{X}}$ or for its components separately [see (13.9, 13.10)]. For the first K EOFS, we find with the help of (13.21).

$$\eta_{\{1\ldots K\}} = 1 - \frac{\epsilon_K}{\mathrm{VAR}\left(\vec{\mathbf{X}}\right)} = 1 - \frac{\sum_{k=K+1}^{m} \lambda_k}{\sum_{k=1}^{m} \lambda_k} = \frac{\sum_{k=1}^{K} \lambda_k}{\sum_{k=1}^{M} \lambda_k} \tag{13.23}$$

If η_j and η_k are the explained variances by two single EOFs $\vec{p}^{\,k}$ and $\vec{p}^{\,j}$ with indices $j > k$ such that $\lambda_j \leq \lambda_k$, then the following inequality holds: $0 < \eta_j \leq \eta_k \leq \eta_{\{j,k\}} \leq 1$, with η_{jk} representing the variance explained by both EOFs $\vec{p}^{\,j}$ and $\vec{p}^{\,k}$).

If the original vector $\vec{\mathbf{X}}$ has m components - what is an adequate truncation K in (13.1)? There is no general answer to this problem which could

also be phrased "Which are the (physically) significant[7] EOFs?" A good answer will depend on the physical problem pursued. One relevant piece of information is the amount of explained variance. One might select K so that the percentage of $\vec{\mathbf{X}}$-variance explained by the first K EOFs, $\eta_{(1\ldots K)}$, passes a certain threshold. Or such that the last kept EOF accounts for a certain minimum variance:

$$\eta_{(1\ldots K)} \leq \kappa_1 < \eta_{(1\ldots K+1)} \qquad \text{or} \qquad \eta_K > \kappa_2 > \eta_{K+1} \tag{13.24}$$

Typical values for κ_1 are 80% or 90% whereas choices of $\kappa_2 = 5\%$ or 1% are often seen.

We have introduced EOFs as patterns which minimize the variance of the residual (13.21). The variance depends on the chosen geometry, and we could replace in (13.21) the square by a *scalar product* $\langle \cdot, \cdot \rangle$ such that

$$\epsilon_K = \mathrm{E}\Big(\langle \vec{\mathbf{X}} - \textstyle\sum_{k=1}^{K} \alpha_k \vec{p}^{\,k}, \vec{\mathbf{X}} - \sum_{k=1}^{K} \alpha_k \vec{p}^{\,k} \rangle \Big) \tag{13.25}$$

The EOF coefficients are then also given as dot products

$$\alpha_k(t) = \langle \vec{\mathbf{X}}_t, \vec{p}^{\,k} \rangle$$

Obviously the result of the analysis depends on the choice of the dot product, which is to some extend arbitrary.

13.3.2 What EOFs are *Not* Designed for ...

There are some words of caution required when dealing with EOFs. These patterns are constructed to represent in an optimal manner *variance* and *covariance* (in the sense of *joint variance*), not *physical connections* or *maximum correlation* (see Chen and Harr, 1993). Therefore they are excellent tools to *compress* data into a few variance-wise significant components. Sometimes people expect more from EOFs, for instance a description of the "coherent structures" (as, for instance, teleconnections). This goal can be achieved only when the data are normalized to variance one, i.e., if the correlation matrix instead of the covariance matrix is considered (see Wallace and Gutzler, 1981). Another expectation is that EOFs would tell us something about the structure of an underlying continuous field from which the data vector $\vec{\mathbf{X}}$ is sampled. Also often EOFs are thought to represent modes of "natural" or "forced" variability. We will discuss these expectations in the following.

[7] Note that the word "significant" used here has nothing to do with "statistical significance" as in the context of testing null hypotheses (see Chapters 8 and 9). Instead, the word "significance" is used in a colloquial manner. We will return to the buzz-word "significance" in Section 13.3.3.

- To demonstrate the limits of an EOF analysis to identify *coherent structures* let us consider the following example with a two-dimensional random vector $\vec{\mathbf{X}} = (\mathbf{X}_1, \mathbf{X}_2)^T$. The covariance matrix $\mathbf{\Sigma}$ and the correlation matrix $\mathbf{\Sigma}'$ of $\vec{\mathbf{X}}$ is assumed to be

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho a \\ \rho a & a^2 \end{pmatrix} \quad \text{and} \quad \mathbf{\Sigma}' = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{13.26}$$

The correlation matrix $\mathbf{\Sigma}'$ is the covariance matrix of the normalized random vector $\vec{\mathbf{X}}' = \mathcal{A}\vec{\mathbf{X}}$ with the diagonal matrix $\mathcal{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1/a \end{pmatrix}$. Obviously both random vectors, $\vec{\mathbf{X}}$ and $\vec{\mathbf{X}}'$, represent the same correlation structure. The relative distribution of variances in the two components $\mathbf{X}_1$ and $\mathbf{X}_2$ depends on the choice of a. Also the eigen-structures of $\mathbf{\Sigma}$ and $\mathbf{\Sigma}'$ differ from each other since the transformation matrix $\mathcal{A}$ is not orthogonal, i.e., it does not satisfy $\mathcal{A}^T = \mathcal{A}^{-1}$.

We will now calculate these eigen-structures for two different standard deviations a of $\mathbf{X}_2$. The eigenvalues of $\mathbf{\Sigma}$ are given by

$$\lambda_{1,2} = \frac{1}{2}\left[1 + a^2 \pm \sqrt{1 - 2a^2 + a^4 + 4(\rho a)^2}\right] \tag{13.27}$$

and the eigenvectors are, apart from proper normalization, given by

$$\vec{p}^{\,1} = \begin{pmatrix} 1 \\ \frac{\lambda_1 - 1}{\rho a} \end{pmatrix} \quad \text{and} \quad \vec{p}^{\,2} = \begin{pmatrix} \frac{1-\lambda_1}{\rho a} \\ 1 \end{pmatrix} \tag{13.28}$$

because of the orthogonality constraint.

In the case of $a^2 \ll 1$ we find

$$\begin{aligned} \lambda &\approx \frac{1}{2}\left[1 + a^2 \pm \sqrt{1 - 2a^2 + 4(\rho a)^2}\right] \\ &\approx \frac{1}{2}\left[1 + a^2 \pm \left(1 - \frac{a^2 - 2(\rho a)^2}{2}\right)\right] = \begin{cases} 1 + (\rho a)^2 \\ a^2(1 - \rho^2) \end{cases} \end{aligned} \tag{13.29}$$

If the two components $\mathbf{X}_1$ and $\mathbf{X}_2$ are perfectly correlated with $\rho = 1$ then the first EOF represents the full variance $\textsc{Var}\left(\vec{\mathbf{X}}\right) = \textsc{Var}(\mathbf{X}_1) + \textsc{Var}(\mathbf{X}_2) = 1 + a^2 = \lambda_1$ and the second EOF represents no variance ($\lambda_2 = 0$). If, on the other hand, the two components are independent, then $\lambda_1 = \textsc{Var}(\mathbf{X}_1) = 1$ and $\lambda_2 = \textsc{Var}(\mathbf{X}_2) = a^2$.

The first EOF is $\vec{p}^{\,1} \approx \begin{pmatrix} 1 \\ \rho a \end{pmatrix} \approx \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, which is reasonable since the first component represents almost all variance in the case of $a^2 \ll 1$. Because of the orthogonality constraint the second EOF is $\vec{p}^{\,2} \approx$

$\begin{pmatrix} -\rho a \\ 1 \end{pmatrix} \approx \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Thus in the case $a \ll 1$ the EOFs are the unit vectors *independently of the size of* ρ.

If we deal with the correlation matrix $\mathbf{\Sigma}'$ the difference of relative importance of the two components is erased. The eigenvalues are given by (13.27) with $a = 1$:

$$\lambda' = \frac{1}{2}\left[2 \pm \sqrt{4\rho^2}\right] = 1 \pm \rho \tag{13.30}$$

The eigenvectors are given by (13.28): If $\rho = 0$ then the eigenvalue (13.30) is double and no unique eigenvectors can be determined. If $\rho > 0$, then the non-normalized EOFs are

$$\vec{p}^{\,1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \vec{p}^{\,2} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{13.31}$$

Because of the positive correlation, the first EOF describes in-phase variations of $\mathbf{X}_1$ and $\mathbf{X}_2$. The orthogonality constraint leaves the second pattern with the representation of the out-of-phase variations.

The "patterns" (13.31) are markedly different from the eigenvectors of the $a \ll 1$-covariance matrix calculated above. Thus, the result of the EOF analyses of two random vectors with the same correlation structure depends strongly on the allocation of the variance within the vector $\vec{\mathbf{X}}$.

This example demonstrates also the impact of using vectors $\vec{\mathbf{X}}$ which carry numbers subjective to different units. If air pressure from midlatitudes is put together with pressure from low latitudes then the EOFs will favor the high-variance midlatitude areas. If a vector is made up of temperatures in units of K and of precipitation in m/sec[8], then patterns of the EOFs will concentrate on the temperature entries.

- An EOF analysis deals with a *vector* of observations $\vec{\mathbf{X}}$. This vector may entertain physically very different entries, as outlined at the beginning of this Section. The EOF analysis does not *know* what type of vector it is analyzing. Instead all *components* of $\vec{\mathbf{X}}$ are considered as equally relevant, independently if they represent a small or a large grid box in case of a longitude $\times$ latitude grid, or a thin or a thick layer (in case of ocean general circulation model output). If we study Scandinavian temperature as given by 10 stations in Denmark and one station in the other Scandinavian states, then the first EOFs will invariably concentrate on Denmark.

  If we deal with a relatively uniform distribution of variance, and if we know that the *characteristic spatial scale* of the considered variable, such

[8]The unit mm/sec is, admittedly, not widely used for precipitation. But precipitation is a rate, often given in mm/day - which is in standard units expressible as m/sec.

as temperature, is comparable to the considered area, then the first EOF will in *most* cases be a pattern with the same sign at all points - simply because of the system's tendency to create anomalies with the same sign in the entire domain. The need to be orthogonal to the first EOF then creates a second EOF with a dipole pattern (which is the largest-scale pattern orthogonal to the uniform-sign first EOF). Our 2-dimensional $(a = 1, \rho > 0)$-case, discussed above, mimics this situation. If, however, the characteristic spatial scale is smaller than the analysis domain then often the first EOF is not a monopole [see, for instance, the SST analysis of Zorita et al. (1992)].

- Do EOFs represent *modes* or *processes* of the physical system from which the data are sampled? In many cases the first EOF may be identified with such a mode or process. For the second and higher indexed EOFs, however, such an association is possible only under very special circumstances (see North, 1984). A severe limitation to this end is the imposed spatial orthogonality of the patterns and the resulting temporal independence of the coefficients (13.22). Thus EOFs can represent only such physical modes which operate independently, and with orthogonal patterns. In most real-world cases, however, processes are interrelated.

13.3.3 Estimating EOFs

The EOFs are parameters of the covariance matrix $\mathbf{\Sigma}$ of a random variable $\vec{\mathbf{X}}$. In practical situations, this covariance matrix $\mathbf{\Sigma}$ is unknown. Therefore, the EOFs have to be estimated from a finite sample $\{\vec{\mathbf{x}}(1) \dots \vec{\mathbf{x}}(n)\}$. In the following, estimations are denoted by $\hat{}$. We assume that the observations represent anomalies, i.e., deviations from the true mean or from the sample mean. However, the analysis can be done in the same way also with data without prior subtraction of the mean.

To *estimate* EOFs from the finite sample $\{\vec{\mathbf{x}}(1) \dots \vec{\mathbf{x}}(n)\}$ two different strategies may be pursued. One strategy considers the finite sample as a finite random variable and calculates orthogonal patterns $\hat{\vec{p}^k}$ which minimize

$$\sum_{l=1}^{n} \left[\vec{\mathbf{x}}(l) - \sum_{k=1}^{K} \hat{\alpha}_k(l) \hat{\vec{p}^k} \right]^2 = \hat{\epsilon}_K \tag{13.32}$$

with coefficients $\hat{\alpha}_k(l) = \sum_{j=1}^{n} \vec{\mathbf{x}}(l)_j \hat{p^k}_j$ given by (13.8).

An alternative approach is via the Theorem of Section 13.2, namely to use the eigenvectors $\hat{\vec{p}^k}$ of the *estimated* covariance matrix

$$\hat{\mathbf{\Sigma}} = \frac{1}{n} \sum_{l=1}^{n} \vec{\mathbf{x}}(l) \vec{\mathbf{x}}(l)^T = \frac{1}{n} \left[\sum_{l=1}^{n} \mathbf{x}_i(l) \mathbf{x}_j(l) \right]_{i,j} \tag{13.33}$$

as *estimators* of the true EOFs $\vec{p}^{\,k}$. Interestingly, both approaches result in the same patterns (H. von Storch and Hannoschöck, 1986).

For the actual computation the following comments might be helpful:

- The samples, which determine the estimated EOFs, enter the procedure only in (13.33) - and in this equation the ordering of the samples is obviously irrelevant. The estimated covariance matrix $\hat{\mathbf{\Sigma}}$ and thus the estimated EOFs, are invariant to the order of the samples.

- When m is the dimension of the analyzed vector $\vec{\mathbf{X}}$ the true covariance matrix $\mathbf{\Sigma}$ as well as the estimated covariance matrix $\hat{\mathbf{\Sigma}}$ have dimension $m \times m$. Therefore the numerical task of *calculating* the eigenvectors and eigenvalues of a sometimes huge $m \times m$ matrix is difficult or even impossible. A wholesale alternative is based on the following little algebraic trick (H. von Storch and Hannoschöck, 1984): *If $\mathcal{Y}$ is a $n \times m$ matrix, then $\mathcal{A} = \mathcal{Y}\mathcal{Y}^T$ and $\mathcal{A}^T = \mathcal{Y}^T\mathcal{Y}$ are $n \times n$- and $m \times m$ matrices which share the same nonzero eigenvalues. If $\mathcal{Y}\vec{q}$ (or $\vec{r}$) is an eigenvector of $\mathcal{A}$ to the eigenvalue $\lambda \neq 0$ then $\vec{q}$ (or $\mathcal{Y}^T\vec{r}$) is an eigenvector of $\mathcal{A}^T$ to the same eigenvalue λ.*

  The estimated covariance matrix $\hat{\mathbf{\Sigma}}$ may be written as $\hat{\mathbf{\Sigma}} = \frac{1}{n}\mathcal{X}\mathcal{X}^T$ with the *data matrix*

$$\mathcal{X} = \begin{pmatrix} x_1(1) & x_1(2) & \dots & x_1(n) \\ x_2(1) & x_2(2) & \dots & x_2(n) \\ \vdots & \vdots & \ddots & \vdots \\ x_m(1) & x_m(2) & \dots & x_m(n) \end{pmatrix} = (\vec{\mathbf{x}}(1)|\dots|\vec{\mathbf{x}}(n)) \tag{13.34}$$

  The n columns of the $n \times m$ data matrix $\mathcal{X}$ are the sample vectors $\vec{\mathbf{x}}(j), j = 1, \dots n$; the rows mark the m coordinates in the original space. The matrix product $\mathcal{X}\mathcal{X}^T$ is a quadratic $m \times m$ matrix even if $\mathcal{X}$ itself is not quadratic. The product $\mathcal{X}^T\mathcal{X}$, on the other hand, is a $n \times n$-matrix. The above mentioned trick tells us that one should calculate the eigenvalues and eigenvectors of the smaller of the two matrices $\mathcal{X}^T\mathcal{X}$ and $\mathcal{X}\mathcal{X}^T$. In practical situations we often have the number of samples n being much smaller than the number of components m.

  A byproduct of the "trick" is the finding that we can estimate only the first n EOFs (or $n-1$ if we have subtracted the overall mean to get anomalies) of the m EOFs of the m-variate random variable.

- Numerically, the EOF analysis of a finite set of observed vectors may by done by a *Singular Value Decomposition* (SVD, see Chapter 14).

$$\mathcal{X}^T = \begin{pmatrix} \tilde{\alpha}_1(1) & \tilde{\alpha}_2(1) & \dots & \tilde{\alpha}_n(1) \\ \tilde{\alpha}_1(2) & \tilde{\alpha}_2(2) & \dots & \tilde{\alpha}_n(2) \\ \vdots & \vdots & & \vdots \\ \tilde{\alpha}_1(n) & \tilde{\alpha}_2(n) & \dots & \tilde{\alpha}_n(n) \end{pmatrix} \mathcal{D}\left(\hat{\vec{p}}^{\,1}|\dots|\hat{\vec{p}}^{\,m}\right)^T \tag{13.35}$$

with a rectangular $n \times m$ matrix $\mathcal{D}$ with zero elements outside the diagonal and positive elements on the diagonal: $d_{ij} = s_i \delta_{ij} \geq 0$. The quadratic $n \times n$ and $m \times m$ matrices to the right and left of $\mathcal{D}$ are orthogonal.

The eigenvalues of the estimated covariance matrix are $\hat{\lambda}_i = s_i^2$. The coefficients of the estimated EOFs are given by $\hat{\alpha}_i = s_i \tilde{\alpha}_i$. Again, there is a maximum of $min(n, m)$ nonzero s_i-values so that at most $min(n, m)$ useful EOFs can be determined.

- The choice of the numerical algorithm is irrelevant for the mathematical character of the product - EOFs are the eigenvectors of the estimated covariance matrix independently if the number crunching has been done via the eigenvector problem or via SVD.

As always, when estimating parameters of a random variable from a finite sample of observations, one may ask how accurate the estimation probably is:

- *Biases*
  If $\hat{\lambda}_k$ is an estimate of the true eigenvalue λ_k and $\hat{\alpha}_k$ the EOF coefficient of the kth estimated EOF the equality of eigenvalues and variance of EOF coefficients is biased (cf. H. von Storch and Hannoschöck, 1986):
  - For the largest eigenvalues λ_k:
    $$\mathrm{E}\left(\hat{\lambda}_k\right) > \lambda_k = \mathrm{VAR}(\alpha_k) > \mathrm{E}(\mathrm{VAR}(\hat{\alpha}_k)) \tag{13.36}$$
  - for the smallest eigenvalues λ_k:
    $$\mathrm{E}\left(\hat{\lambda}_k\right) < \lambda_k = \mathrm{VAR}(\alpha_k) < \mathrm{E}(\mathrm{VAR}(\hat{\alpha}_k)) \tag{13.37}$$

  The relation (13.36,13.37) means that the large (small) eigenvalues are systematically over-(under)estimated, and that the variance of the random variable $\hat{\alpha}_k = \langle \vec{\mathbf{X}}, \hat{\vec{p}^k} \rangle$ which are expansion coefficients when projecting $\vec{\mathbf{X}}$ on the random variable "estimated EOFs" is systematically over- or underestimated by the sample variance $\widehat{\mathrm{VAR}(\hat{\alpha}_k)} = \hat{\lambda}_k$ derived from the sample $\{\vec{\mathbf{x}}(1) \ldots \vec{\mathbf{x}}(n)\}$. Similarly, $\mathrm{COV}(\hat{\alpha}_k, \hat{\alpha}_j) \neq \widehat{\mathrm{COV}(\hat{\alpha}_k, \hat{\alpha}_j)} = 0$.

- *"Selection Rules"*
  So-called *selection rules* have been proposed. One often used is named "Rule N"(Preisendorfer and Overland, 1982), which is supposed to determine the physically "significant" EOFs. The basic concept is that the full phase space is the sum of a subset in which all variations are purely noise and of a subset whose variability is given by dynamical processes. The signal-subspace is spanned by well-defined EOFs whereas in the noise-subspace no preferred directions exist. For the eigenvalue-spectrum this

assumption implies that the eigenvalues of the EOFs spanning the signal-subspace are unequal and that the eigenvalues in the noise-subspace are all identical.

The selection rules compare the distributions of sample eigenvalue-spectra, representative for the situation that all or the $m - K$ smallest true eigenvalues (K being specified a-priori or determined recursively) are all alike, with the actually observed sample eigenvalue spectrum. All those estimated eigenvalues which are larger than the, say, 95%-percentile of the (marginal) distribution of the reference "noise spectra", are selected as *significant* at the 5%-level.

The problem with this approach is that this selection rule is claimed to be a *statistical test* which supposedly is capable of accepting, with a given risk, the alternative hypothesis that all EOFs with an index smaller than some number $m-K$ represent "signals" of the analyzed data field. The null hypothesis tested would be "all eigenvalues are equal", and the rejection of this null hypothesis would be the acceptance of the alternative "not all eigenvalues are equal". The connection between this alternative and the determination of a "signal subspace" is vague. Also the above sketched approach does not consider the quality of the estimation of the patterns; instead the selection rules are concerned with the eigenvalues only.

I recommend forgetting about the identification of "significant" EOFs by means of selection rules and resorting to more honest approaches like *North's rule-of-thumb* outlined in the next paragraph.

- *North's Rule-of-Thumb*
  Using a scale argument North et al. (1982) found as the "typical errors"

$$\Delta\lambda_k \approx \sqrt{\frac{2}{n}}\lambda_k \tag{13.38}$$

$$\Delta\hat{\vec{p}}^k \sim \frac{\Delta\lambda_k}{\lambda_j - \lambda_k}\vec{p}^{\,j} \tag{13.39}$$

  with λ_j being the eigenvalue closest to λ_i and n being the number of *independent* samples. Approximation (13.39) compares patterns, and not the lengths of vectors, since we are dealing with normalized vectors.

  - The first order error $\Delta\hat{\vec{p}}^k$ is of the order of $\sqrt{\frac{1}{n}}$. The convergence to zero is slow.
  - The first order error $\Delta\hat{\vec{p}}^k$ is orthogonal to the true EOF $\vec{p}^{\,k}$.
  - The estimation of the EOF $\vec{p}^{\,k}$ is most contaminated by the patterns of those other EOFs $\vec{p}^{\,j}$ which belong to eigenvalues λ_j closest to λ_k. The contamination will be the more severe the smaller the difference $\lambda_j - \lambda_k$ is.

North et al. (1982) finally formulated the following "rule-of-thumb":

> "If the sampling error of a particular eigenvalue $\Delta\lambda$ is comparable or larger than the spacing between λ and a neighboring eigenvalue, then the sampling error $\Delta\vec{p}$ of the EOF will be comparable to the size of the neighboring EOF."

When using this rule-of-thumb one should be careful not to oversee the condition of independent samples - in most geophysical data this assumption is not valid.

13.3.4 Example: Central European Temperature

As an example we consider the covariance structure of the winter mean temperature anomalies (i.e., deviations from the overall winter mean) at eleven Central European stations (Werner and H. von Storch, 1993). Thus $m = 11$. Relatively homogeneous time series were available for eighty winters from 1901 to 1980. For both of the 40-year interval before and after 1940 an EOF analysis was performed. The results of the analysis are very similar in the two intervals - as a demonstration we show in Figure 13.3 the first two EOFs for both periods. The representation of the patterns deviates from the definition introduced above: The contribution of the k-th EOF to the full signal is given by $\alpha_k(t)\vec{p}^{\,k}$. According to our definitions the variance of the coefficient is given by the k-th eigenvalue λ_k and the vector has unit length. For a better display of the results sometimes a different normalization is convenient, namely $\alpha_k\vec{p}^{\,k} = (\alpha_k/\sqrt{\lambda_k}) \times (\vec{p}^{\,k}\sqrt{\lambda_k}) = \alpha'_k\vec{p}^{\,'k}$. In this normalization the coefficient time series has variance one for all indices k and the relative strength of the signal is in the patterns $\vec{p}^{\,'}$. A typical coefficient is $\alpha' = 1$ so that the typical reconstructed signal is $\vec{p}^{\,'}$. In this format the first two EOFs and their time coefficients obtained in the analysis of Central European winter temperature are shown in Figures 13.3 and 13.4.

In both time periods the first EOF has a positive sign at all locations, represents about 90% of the total variance and exhibits "typical anomalies" of the order of $1 - 2K$. The second EOF represents a northeast-southwest gradient, with typical anomalies of $\pm 0.5K$, and accounts for 6% and 7% of the variance in the two time periods. The remaining 9 EOFs are left to represent together the variance of mere 5%.

In Figure 13.4 the EOF coefficients are shown. As mentioned above, they are normalized to variance one. The first coefficient $\alpha_1(t)$ varies most of the time between ± 1 but exhibits a number of spiky excursions to large negative values < -2. Together with the information provided by the patterns (Figure 13.3) such large negative coefficients represent extremely cold winters, such as 1940 and 1941, with mean negative anomalies of the order of $< -4K$. The distribution of the first EOF coefficient is markedly skewed (Figure 13.5) whereas the distribution of the second coefficient is fairly symmetric. The

time series of the 2nd coefficient, depicted in Figure 13.4 shows no dramatic outliers but, interestingly, an upward trend translates at the stations with a slow warming of the Alpine region ($\approx 0.005K/yr$) and a gradual cooling ($\approx 0.01K/yr$) in the lowlands.

This is about all that the EOFs can tell us about the evolution of winter mean temperature in Central Europe in the years 1901-80. We will come back to this example in Section 13.4.4.

13.4 Canonical Correlation Analysis

We assume for this section again that the expectations of the considered random vectors $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ vanish: $\vec{\mu}_X = \vec{\mu}_Y = 0$.

13.4.1 Definition of Canonical Correlation Patterns

In the Canonical Correlation Analysis [CCA, proposed by Hotelling (1936) and introduced into climate research by, among others, Barnett and Preisendorfer (1987)] not one random vector $\vec{\mathbf{X}}$ is expanded into a finite set of

vectors but a pair of two simultaneously observed vectors $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$:

$$\vec{\mathbf{X}}_t = \sum_{k=1}^{K} \alpha_k^X(t)\vec{p}_X^k \quad \text{and} \quad \vec{\mathbf{Y}}_t = \sum_{k=1}^{K} \alpha_k^Y(t)\vec{p}_Y^k \tag{13.40}$$

with the same number K. The dimensions m_X and m_Y of the vectors $\vec{p}_X^k$ and $\vec{p}_Y^k$ will in general be different. The expansion is done in such a manner that

1. The coefficients $\alpha_k^X(t)$ and $\alpha_k^Y(t)$ in (13.40) are *optimal* in a least square sense [i.e., for given patterns $\vec{p}_X^k$ and $\vec{p}_Y^k$ the norms $\| \vec{\mathbf{X}}_t - \sum_{k=1}^K \alpha_k^X(t)\vec{p}_X^k \|$ and $\| \vec{\mathbf{Y}}_t - \sum_{k=1}^K \alpha_k^Y(t)\vec{p}_Y^k \|$ are minimized, as in (13.2)]. This condition implies (see Section 13.2.1) that

   $$\alpha_k^X = \langle (\vec{p}_X^k)_A, \vec{\mathbf{X}} \rangle \quad \text{and} \quad \alpha_k^Y = \langle (\vec{p}_Y^k)_A, \vec{\mathbf{Y}} \rangle \tag{13.41}$$

   with certain *adjoint patterns* $(\vec{p}_X^k)_A$ and $(\vec{p}_Y^k)_A$ given by (13.6).

2. The correlations

   - between α_k^X and α_l^X
   - between α_k^Y and α_l^Y
   - between α_k^X and α_l^Y

   are zero for all $k \neq l$.

3. The correlation between α_1^X and α_1^Y is maximum.

4. The correlation between α_2^X and α_2^Y is the maximum under the constraints of 2) and 3). The correlations for the higher indexed pairs of coefficients satisfy similar constraints (namely of being maximum while being independent with all previously determined coefficients.)

It can be shown (see, for instance, Zorita et al., 1992) that the adjoint patterns are the eigenvectors of somewhat complicated looking matrices, namely:

$$\mathcal{A}_X = \mathbf{\Sigma}_X^{-1}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1}\mathbf{\Sigma}_{XY}^T \quad \text{and} \quad \mathcal{A}_Y = \mathbf{\Sigma}_Y^{-1}\mathbf{\Sigma}_{XY}^T\mathbf{\Sigma}_X^{-1}\mathbf{\Sigma}_{XY} \tag{13.42}$$

Here $\mathbf{\Sigma}_X$ and $\mathbf{\Sigma}_Y$ are the covariance matrices of $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$. $\mathbf{\Sigma}_{XY}$ is the cross-covariance matrix of $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$, i.e., $\mathbf{\Sigma}_{XY} = \mathrm{E}\left(\vec{\mathbf{X}}\vec{\mathbf{Y}}^T\right)$ if $\mathrm{E}\left(\vec{\mathbf{X}}\right) = \mathrm{E}\left(\vec{\mathbf{Y}}\right) = 0$. The matrix $\mathcal{A}_X$ is a $m_X \times m_X$ matrix and $\mathcal{A}_Y$ is a $m_Y \times m_Y$ matrix. The two matrices $\mathcal{A}_X$ and $\mathcal{A}_Y$ may be written as products $\mathcal{B}_1\mathcal{B}_2$ and $\mathcal{B}_2\mathcal{B}_1$ with two matrices $\mathcal{B}_1$ and $\mathcal{B}_2$. Therefore the two matrices share the same nonzero eigenvalues, and if $\vec{p}_X^k$ is an eigenvector of $\mathcal{A}_X$ with an eigenvalue $\lambda \neq 0$ then $\mathbf{\Sigma}_Y^{-1}\mathbf{\Sigma}_{XY}^T\vec{p}_X^k$ is an eigenvector of $\mathcal{A}_Y$ with the same eigenvalue.

Note that for univariate random variables $\vec{\mathbf{X}} = \mathbf{X}$ and $\vec{\mathbf{Y}} = \mathbf{Y}$ the two matrices $\mathcal{A}_X$ and $\mathcal{A}_Y$ in (13.42) reduce to the squared correlations between $\mathbf{X}$ and $\mathbf{Y}$.

The k-adjoint pattern is given by the eigenvector with the k-largest eigenvalue of $\mathcal{A}$. The correlation between α_k^X and α_k^Y is given by the k-th largest nonzero eigenvalue of $\mathcal{A}_X$ or $\mathcal{A}_Y$.

The covariance between the "Canonical Correlation Coefficients" α_k^X and the original vector $\vec{\mathbf{X}}$ is given by

$$\mathrm{E}\Big(\alpha_k^X \vec{\mathbf{X}}\Big) = \mathrm{E}\big(\alpha_k^X \textstyle\sum_i \alpha_i^X \vec{p}_X^{\,i}\big) = \vec{p}_X^{\,k} \tag{13.43}$$

so that, because of $\alpha_k^X = \vec{\mathbf{X}}^T (\vec{p}_X^{\,k})_A$:

$$\vec{p}_X^{\,k} = \mathbf{\Sigma}_X (\vec{p}_X^{\,k})_A \quad \text{and} \qquad \vec{p}_Y^{\,k} = \mathbf{\Sigma}_Y (\vec{p}_Y^{\,k})_A \tag{13.44}$$

Thus, to determine the "Canonical Correlation Patterns" (CCP) and the canonical correlation coefficients one has first to calculate the covariance matrices and cross covariance matrices. From products of these matrices (13.42) the adjoint patterns are derived as eigenvectors. With the adjoint pattern the CCPs are calculated via (13.44) and the coefficients through (13.41). Because of the specific form of the matrices, it is advisable to solve the eigenvector problem for the smaller one of the two matrices.

13.4.2 CCA in EOF Coordinates

A simplification of the mathematics may be obtained by first transforming the random vectors $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ into a low-dimensional EOF-space, i.e., by expanding

$$\vec{\mathbf{X}} \approx \vec{\mathbf{X}}^S = \sum_{i=1}^{K} (\beta_i^X)(\sqrt{\nu_i^X}\,\vec{e}_X^{\,i}) \quad \text{and} \quad \vec{\mathbf{Y}} \approx \vec{\mathbf{Y}}^S = \sum_{i=1}^{K} (\beta_i^Y)(\sqrt{\nu_i^Y}\,\vec{e}_Y^{\,i}) \tag{13.45}$$

with EOFs $\vec{e}_X^{\,i}$ of $\vec{\mathbf{X}}$ and $\vec{e}_Y^{\,i}$ of $\vec{\mathbf{Y}}$. The numbers ν_i^X and ν_i^Y, which are the eigenvalues associated with the EOFs, are introduced to enforce $\textsc{Var}\big(\beta_i^X\big) = \textsc{Var}\big(\beta_i^Y\big) = 1$. Equations (13.45) may be written more compactly with the help of matrices $\mathcal{E} = (\vec{e}^{\,1} | \ldots | \vec{e}^{\,K})$ with the EOFs in their columns (so that $\mathcal{E}\mathcal{E}^T = 1$ and $\mathcal{E}^T\mathcal{E} = 1$) and diagonal matrices $\mathcal{S} = (diag\sqrt{\nu_i})$:

$$\vec{\mathbf{X}}^S = \mathcal{E}_X \mathcal{S}_X \vec{\beta}^X \quad \text{and} \quad \vec{\mathbf{Y}}^S = \mathcal{E}_Y \mathcal{S}_Y \vec{\beta}^Y \tag{13.46}$$

When we operate with objects in the EOF coordinates we add a tilde $\tilde{}$. In these coordinates we have $\tilde{\mathbf{\Sigma}}_X = 1$ and $\tilde{\mathbf{\Sigma}}_Y = 1$ and the CCA matrices (13.42) are of the simpler and symmetric form

$$\tilde{\mathcal{A}}_X = \widetilde{\mathbf{\Sigma}_{XY}}\,\widetilde{\mathbf{\Sigma}_{XY}}^{\,T} \quad \text{and} \quad \tilde{\mathcal{A}}_Y = \widetilde{\mathbf{\Sigma}_{XY}}^{\,T}\,\widetilde{\mathbf{\Sigma}_{XY}} \tag{13.47}$$

In the EOF coordinates the CCA patterns are orthogonal, so that *in these coordinates* $\widetilde{\vec{p}_X^k} = (\widetilde{\vec{p}_X^k})_A$. The procedure to get the CC patterns and the adjoints in the original Euclidean-space is the same for $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$, so that we consider only the $\vec{\mathbf{X}}$-case and drop the index "X" as well as the index "i" in the following for convenience. Also we identify the full representation $\vec{\mathbf{X}}$ with the truncated presentation $\vec{\mathbf{X}}^S$.

- The CCA coefficients α at any given time should be independent of the coordinates. Thus, if $\vec{\beta}(t)$ is the state of $\vec{\mathbf{X}}$ in the EOF coordinates (13.45) and $\vec{x}(t)$ in the original Euclidean coordinates, then the CCA coefficients shall be given as the dot product of this vector of state with adjoint patterns $\vec{p}_A$ and $\widetilde{\vec{p}_A}$:

$$\alpha(t) = \widetilde{\vec{p}_A}^T \vec{\beta} = \vec{p}_A{}^T \vec{x}(t) \tag{13.48}$$

- The initial transformation (13.45), $\vec{x} = \mathcal{E}\mathcal{S}\vec{\beta}$, describes the transformation of the CC patterns from the EOF coordinates to the Euclidean coordinates:

$$\vec{p}^{\,k} = \mathcal{E}\mathcal{S}\widetilde{\vec{p}^{\,k}} \tag{13.49}$$

- To get the back transformation of the adjoints we insert (13.45) into (13.48) and get $\vec{p}_A{}^T\vec{x} = \widetilde{\vec{p}_A}^T\vec{\beta} = \widetilde{\vec{p}_A}^T\mathcal{S}^{-1}\mathcal{E}^T\vec{x}$ and

$$\vec{p}_A^{\,k} = \mathcal{E}\mathcal{S}^{-1}\widetilde{\vec{p}_A^{\,k}} \tag{13.50}$$

In general we have $\mathcal{S} \neq \mathcal{S}^{-1}$ so that neither the property "self adjoint", i.e, $\widetilde{\vec{p}_A^{\,k}} = \widetilde{\vec{p}^{\,k}}$, nor the property "orthogonal", i.e. $\widetilde{\vec{p}^{\,k}}^T\widetilde{\vec{p}^{\,l}} = 0$ if $k \neq l$, are valid after the back transformation into the Euclidean space.

In the EOF coordinates we can establish a connection to the EOF calculus (Vautard; pers. communication). For convenience we drop now the ˜ marking objects given in the EOF coordinates. First we concatenate the two vectors $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ to one vector $\vec{\mathbf{Z}} = (\vec{\mathbf{X}}, \vec{\mathbf{Y}})$ and calculate the EOFs $\vec{e}^{\,i}$ of this new random vector. These EOFs are the eigenvectors of the joint covariance matrix

$$\mathbf{\Sigma}_Z = \begin{pmatrix} \mathbf{\Sigma}_X & \mathbf{\Sigma}_{XY} \\ \mathbf{\Sigma}_{XY}^T & \mathbf{\Sigma}_Y \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{\Sigma}_{XY} \\ \mathbf{\Sigma}_{XY}^T & 1 \end{pmatrix} \tag{13.51}$$

A vector $\vec{p} = (\vec{p}_X, \vec{p}_Y)$ is an eigenvector of $\mathbf{\Sigma}_Z$ if

$$\frac{1}{\lambda - 1}\mathbf{\Sigma}_{XY}\vec{p}_Y = \vec{p}_X \quad \text{and} \quad \frac{1}{\lambda - 1}\mathbf{\Sigma}_{XY}^T\vec{p}_X = \vec{p}_Y \tag{13.52}$$

so that $\vec{p}_X$ and $\vec{p}_Y$ have to satisfy

$$\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_{XY}^T\vec{p}_X = (\lambda - 1)^2\vec{p}_X \quad \text{and} \quad \mathbf{\Sigma}_{XY}^T\mathbf{\Sigma}_{XY}\vec{p}_Y = (\lambda - 1)^2\vec{p}_Y \tag{13.53}$$

Thus the two components $\vec{p}_X$ and $\vec{p}_Y$ of the joint "extended" EOF of $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ form a pair of canonical correlation patterns of $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$. Note that this statement depends crucially on the nontrivial assumption $\mathbf{\Sigma}_X = \mathbf{\Sigma}_Y = 1$.

13.4.3 Estimation: CCA of Finite Samples

The estimation of CC patterns, adjoints and CC coefficients is made in a straightforward manner by estimating the required matrices $\mathbf{\Sigma}_X$, $\mathbf{\Sigma}_Y$ and $\mathbf{\Sigma}_{XY}$ in the conventional way [as in (13.33)], and multiply the matrices to get estimates $\widehat{\mathcal{A}_X}$ and $\widehat{\mathcal{A}_Y}$ of $\mathcal{A}_X$ and $\mathcal{A}_Y$. The calculation is simplified if the data are first transformed (13.45) into EOF coordinates.

In the case of EOFs we had seen that the first eigenvalues of the estimated covariance matrix *over*estimate the variance which is accounted for by the first EOFs. This overestimation makes sense if one considers the fact that the EOFs must represent a certain amount of variance of the full (*infinite*) random variable $\vec{\mathbf{X}}$, whereas the estimated EOF represents a fraction of variance in the *finite* sub-space given by the samples. In the case of the CCA we have a similar problem: The correlations are *over*estimated - since they are fitted to describe similar behavior only in a finite subspace, given by the samples, of the infinite space of possible random realizations of $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$. This overestimation decreases with increasing sample size and increases with the number of EOFs used in the a-priori compression (13.45).

Zwiers (1993) discusses the problem of estimating the correlations. Following Glynn and Muirhad (1978) he proposes to improve the straightforward estimator $\hat{\rho}_k$ by using the formula

$$\hat{\theta}_k = Z_k - \frac{1}{2n\hat{\rho}_k}\left[m_X + m_y - 2 + \hat{\rho}_k^2 + 2(1-\hat{\rho}_k^2)\sum_{j\neq k}\frac{\hat{\rho}_j^2}{\hat{\rho}_k^2 - \hat{\rho}_j^2}\right] \tag{13.54}$$

with $Z_k = tanh^{-1}(\hat{\rho}_k)$ and $\theta = tanh^{-1}(\rho_k)$. Glynn and Muirhead show that $\hat{\theta}_k$ is an unbiased estimator of θ_k up to $\mathcal{O}(n^{-2})$ and that $\textsc{Var}\left(\hat{\theta}\right) = \frac{1}{n}$ up to order $\mathcal{O}(n^{-2})$. An approximate 95% confidence interval for the $k-$th canonical correlation is then given by $tanh(\hat{\theta}_k) \pm \frac{2}{\sqrt{n}}$. However, Zwiers (1993) found in Monte Carlo experiments that the correction (13.54) represents an improvement over the straightforward approach but that substantial bias remains when the sample size is small.

13.4.4 Example: Central European Temperature

We return now to the Central European temperature in winter, with which we have dealt already in Section 13.3.4. Werner and H. von Storch (1993) analysed simultaneously the large-scale circulation, as given by the seasonal mean sea-level air-pressure (SLP) anomaly over the North Atlantic and Europe, and the Central European temperature given at $m_T = 11$ locations.

The objective of this exercise was to determine to what extent the regional temperature is controlled by large-scale circulation anomalies.[9]

With first 40 years of the full 1901-1980 data set the CCA was done. The data were first projected on EOFs as given by (13.45). The number of EOFs retained was determined in such a way that an increase by one EOF would change the canonical correlations only little. The first pair of patterns goes with a correlation of 70%. The patterns (Figure 13.6), which are plotted here as a "typical simultaneously appearing pair of pattern" (by normalizing the canonical correlation coefficients to variance one) indicate a simple physically plausible mechanism: In winters with a persistent anomalous southwesterly flow, anomalous warm maritime air is advected into Europe so that the winter mean temperatures are everywhere in central Europe above normal, with typical anomalies of the $1^o - 2^oC$. If, however, an anomalous high pressure system is placed south of Iceland, then the climatological transport of maritime air into Central Europe is less efficient, and temperature drops by one or two degrees.

With the full data set, from 1901-80 the correlation of the coefficients α_1^T and α_1^{SLP} is recalculated and found to be only 0.64 compared to 0.70 derived from the "fitting" interval 1901-40. Also the percentage η of temperature variance at the eleven Central European locations accounted for by the temperature-pattern $\vec{p}_T^1$ is computed from the full data set. The results, shown as lower numbers in the top panel of Figure 13.6, vary between 22% (at the northernmost station Fanø) to 58% (at the northern slope of the Alps, at station Hohenpeissenberg).

With two CCA pairs a "downscaling model" was designed to estimate the temperature variations only from the SLP variations without any local information. The result of this exercise is shown for the station Hamburg in Figure 13.7: The year-to-year variations are nicely reproduced - and more than 60% of the 80-year variance is represented by the statistical model - but the long-term (downward) trend is captured only partly by the CCA model: the SLP variations allude to a decrease by -2.6^oC whereas the real decrease has only been -1.0^oC.

Finally, the output of a climate GCM has been analysed whether it reproduces the connection represented by the CCA-pairs. The regional temperature from a GCM output is given at grid points (the proper interpretation of what these grid points represent is not clear). Therefore the 11 Central European stations are replaced by 6 grid points. The first pair of CC patterns, derived from 100 years of simulated data, is shown in Figure 13.8. The patterns are similar to the patterns derived from observed data, with an anomalous southwesterly flow being associated with an overall warming

[9] In any kind of correlation-based analysis a positive result (i.e., the identification of a high correlation) is no proof that the two considered time series are connected through a cause-effect relationship. It can very well be that a third (unknown) is controlling both analysed time series. In the present case the physical concept that the large-scale circulation affects regional temperatures is invoked to allow for the interpretation "control".

of the order of one to two degrees. The details, however, do not fit. First, the correlation is only 0.53 compared to 0.64 in the real world. Second, the structure within Central Europe is not reproduced: Maximum temperature anomalies are at the westernmost grid points and minimum values at the easternmost. The local explained variances η are much higher for the GCM output (with a maximum of 96% and a minimum of 50% compared to 58% and 22% in the real world).

13.5 Optimal Regression Patterns

In the past years, two new techniques for the derivation of spatial patterns have been introduced. Both techniques are based on the idea of optimizing linear regression equations. In the following we sketch both approaches.

13.5.1 Empirical Orthogonal Teleconnections

In the this technique, proposed by van den Dool et al. (2000), a random vector $\vec{\mathbf{X}}_t$ is expanded into a series (13.1): $\vec{\mathbf{X}}_t = \sum_{k=1}^{K} \alpha_k(t)\vec{p}^{\,k} + \vec{n}_t$. The patterns, named *Orthogonal Empirical Teleconnections* (EOTs), are determined sequentially.

To do so, we label the random vector $\vec{\mathbf{X}}_t$ as $\vec{\mathbf{X}}_t^{(0)}$ and its components as as $\mathbf{X}_{i;t}^{(0)}$. Then regression equations are derived, mapping the time series $\mathbf{X}_{i;t}^{(0)}$ at a component i on all other components. If the regression from component i on component j is called $r_{i;j}^{(0)}$ then the total proportion of variance explained by these regressions, given by

$$V_i^{(0)} = \sum_j \mathrm{VAR}\Big(r_{i;j}^{(0)}\mathbf{X}_{i;t}^{(0)} - \mathbf{X}_{j;t}^{(0)}\Big) \tag{13.55}$$

measures how representative the point i is for the full vector. As first EOT is chosen the vector of regression coefficients $r_{i;j}^{(0)}$ with the index i_1 with maximum V_{i_1}. That is $\vec{p}^{\,1} = (r_{i_1;j}^{(0)})_j$, and $\alpha_1(t) = \mathbf{X}_{i_1;t}^{(0)}$.

In a second step, and in all further steps, the "already explained" part is subtracted, i.e.,

$$\vec{\mathbf{X}}_t^{(1)} = \vec{\mathbf{X}}_t^{(0)} - \alpha_1(t)\vec{p}^{\,1} \tag{13.56}$$

is formed, and again regression equations for all points calculated. The second pattern $\vec{p}^{\,2}$ is then the vector of regression coefficients $r_{i;j}^{(1)}$ with maximum $V_i^{(1)}$. Of course, $r_{i_2;i_1}^{(1)} = 0$ and $r_{i_2;i_2}^{(1)} = 1$. The time series $\alpha_2(t)$ is the reduced component time series i_2, i.e., $\alpha_2(t) = \mathbf{X}_{i_2;t}^{(1)}$. Because of the construction with regression equations and the subtraction (13.56), the coefficient time series are uncorrelated: $\mathrm{COV}(\alpha_1, \alpha_2) = 0$.

In this way, a series of pattern $\vec{p}$ are constructed, with uncorrelated coefficient time series and specified skill in specifying the random vector as a whole. Since the coefficients are uncorrelated, the total variance of $\vec{\mathbf{X}}$ is the sum of the variances of the α's. When the data $\vec{\mathbf{X}}$ can be displayed as a map, then also the EOTs may be represented as maps.

In a number of cases, presented by van den Dool et al. (2000), the EOTs were almost as good in specifying variance as EOFs were. Van den Dool et al. (2000) note a number of advantages of their approach, among others that the patters are linked to specific geographical points (when $\vec{\mathbf{X}}$ is made up of spatially distributed variables). In that sense, EOTs are automatically "regionalized" and may therefore serve as an alternative to Rotated EOFs. They may help to identify "important" sampling locations. For instance, when analysing January mean temperature of the contiguous US, they found as first two EOTs a location in West Kentucky, representing 38% of the total variance, and at another location in East Wyoming, representing 31%.

An interesting variant of this method concerns the choice of the indices i_1 etc. There is no need to maximize $V_i^{(0)}$; instead a component could be chosen independently of the variance for dynamical reasons, for instead a center of action of a teleconnection pattern.

Another variant is to reverse the role of space and time; then the EOTs are regression patterns relating the same point at some time with itself for all other times. The series α is then a spatial distribution at a certain fixed time. In the example mentioned above, van den Dool et al. (2000) identified Januaries 1977 and 1937 as such "characteristic" times.

13.5.2 Redundancy Analysis

Redundancy Analysis (RDA) is a variant of Canonical Correlation Analysis. CCA determines in a pair of random vectors patters whose coefficients share a maximum of correlation,; thus for CCA the two random vectors may be swapped without changing the result. In contrast, in RDA the two random vectors $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ play different roles, with one, say $\vec{\mathbf{X}}$, being a predictor, and the other, say $\vec{\mathbf{Y}}$, being the predictand.[10] Patterns are determined so that a maximum proportion of variance of $\vec{\mathbf{Y}}$ is accounted for by regressing "the $\vec{\mathbf{X}}$-patterns" on $\vec{\mathbf{Y}}$.

The concept was invented by Tyler (1982), and worked out for climate applications by H. von Storch and Zwiers (1999). Since the mathematics are a bit involved, we limit ourselves here to a short heuristic introduction.

We begin with the same formalism as in case of CCA, i.e., with expansion (13.40). For a given number K RDA determines a set of $\vec{\mathbf{X}}$-patterns $\vec{p}_X^j$, $j = 1 \ldots k$, so that the regression of the time coefficients α_k^X on $\vec{\mathbf{Y}}$ is maximum.

[10] The use of the conventional expressions "predictor" and "predictand" does not imply that forecasts in time are made; indeed the terms "specificator" and "specificand" would be more precise.

This regression maps the pattern $\vec{p}_X^{\,j}$ on the $\vec{\mathbf{Y}}$-pattern $\vec{p}_Y^{\,j}$. These pattern can be chosen so that the $\vec{\mathbf{Y}}$-patterns form a set of orthonormal patterns, whereas the $\vec{\mathbf{X}}$-patterns are linearly independent.

The $\vec{\mathbf{X}}$-patterns are invariant under non-singular transformations; this is meaningful as the regression should not depend on the units of the predictor; however, since the $\vec{p}_Y^{\,j}$-patterns are constructed for an optimal representation of $\vec{\mathbf{Y}}$-variance, these patterns are invariant only under orthonormal transformation - since the variance would otherwise be changed.

The patterns are given as solutions of eigen-problems, namely

$$\left[\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yx}\right]\vec{p}_X^{\,j} = \lambda_j \vec{p}_X^{\,j} \tag{13.57}$$

$$\left[\mathbf{\Sigma}_{yx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xy}\right]\vec{p}_Y^{\,j} = \lambda_j \vec{p}_Y^{\,j} \tag{13.58}$$

The eigenvalues λ_j represent the proportion of $\vec{\mathbf{Y}}$-variance accounted for through the regression from $\vec{p}_X^{\,j}$ on $\vec{p}_Y^{\,j}$.

The technique has been used for the purpose of empirical downscaling, for instance for estimating wave height statistics from monthly mean air pressure distributions (WASA, 1998).