

4.26.2.6 To appear: *Encyclopedia of Life Support Systems*, EOLSS Publishers Co. Ltd.

SPATIO-TEMPORAL METHODS IN CLIMATOLOGY

Christopher K. Wikle, Department of Statistics, University of Missouri-Columbia, USA

Keywords: Data assimilation, Empirical orthogonal function, Principal oscillation pattern, Canonical correlation

Contents

1. Introduction
 2. Descriptive Statistical Methods
 - 2.1 Empirical Orthogonal Function (EOF) Analysis
 - 2.2 Principal Oscillation Pattern (POP) Analysis
 - 2.3 Space-Time Canonical Correlation Analysis (CCA)
 - 2.4 Space-Time Spectral Analysis
 3. Future Directions
- Bibliography

Glossary

AR(1): Autoregressive model of order one. The present state of a system can be described as a linear function of the state at the previous time, plus time-independent noise.

data assimilation: Method of optimally combining irregularly spaced observations with dynamical constraints to produce dynamically consistent fields on regular grids.

CCA: Canonical Correlation Analysis

EOF: Empirical Orthogonal Function

MAF: Minimum/maximum Autocorrelation Factor; A method for obtaining the optimal eigenfunctions to maximize the autocorrelation in a vector AR(1) process.

MOM: Method of Moments; Statistical estimation approach that seeks estimators by equating empirical and theoretical moments.

normal mode: Natural modes of variability (or oscillations) of a physical system.

PCA: Principal Component Analysis

POP: Principal Oscillation Pattern Analysis

Summary

This article contains a comprehensive discussion of descriptive spatio-temporal statistical methods that have been applied extensively in the climatological sciences. The climate system is composed of many processes that exhibit complicated variability over a vast range of spatial and temporal scales. Data sets of measurements collected on this system are typically very large by statistical standards, and their analysis typically requires dimension reduction in space and/or time. Scientists have developed or borrowed and refined many descriptive statistical techniques that aid in the summary and interpretation of these data. The focus here is on a subset of some of the most useful methodologies: empirical orthogonal function (EOF) analysis, principal oscillation pattern (POP) analysis, spatio-temporal canonical correlation analysis (CCA), and space-time spectral analysis. These methods are described in detail along with physical motivation, discussion of estimation issues, and practical considerations.

1 Introduction

The climate system can be described as the superposition of a set of deterministic, multivariate, and nonlinear interactions over an enormous range of spatial and temporal scales. In order to understand this system, scientists must observe, summarize, make inference, and ultimately predict its behavior at each scale of variability, as well as the interaction between these scales. Unfortunately, although the system is deterministic in principle, the collective knowledge is incomplete at each of the observation, summarization, and inference stages, and thus ultimate understanding is clouded by uncertainty. Consequently, by the time one considers the prediction phase, this lack of certainty, combined with the nonlinear dynamics of the system, contributes to what is now known as dynamical chaos. Although one is always faced with the inherent chaotic nature of the climate system, many of the relevant scientific questions can be approached from

a probabilistic viewpoint, which allows useful inference to be made in the presence of uncertainty, at least for relatively large spatial scales and relatively short temporal scales. Furthermore, one is then able to look for possible associations within and between variables in the system, which may ultimately extend the still incomplete physical theory.

Central to the observation, summarization, inference, and prediction of the atmosphere/ocean system is *data*. Unfortunately, all data come bundled with error. This is an inescapable fact of scientific life. In particular, along with the obvious errors associated with the measuring, manipulating, and archiving of data, there are errors due to the discrete spatial and temporal sampling of an inherently continuous system. Consequently, there are always scales of variability that are unresolvable, and which will surely contaminate the observations. In atmospheric science, this is considered a form of “turbulence”, and corresponds to the well-known aliasing problem in time-series analysis and the “nugget effect” in geostatistics. Furthermore, atmospheric and oceanic data are rarely sampled at spatial or temporal locations that are optimal for the solution of a specific scientific problem. For instance, there is an obvious bias in data coverage towards areas where population density is large and, due to the cost of obtaining observations, towards a country whose Gross Domestic Product is relatively large. Thus, the location of a measuring site and its temporal sampling frequency may have very little to do with science. To gain scientific insight, these uncertainties must be considered when framing scientific questions, choosing analysis techniques, and interpreting results. This task is complicated further since atmospheric and oceanic data are nearly always correlated in space and time. In this case most of the traditional statistical methods taught in introductory statistics courses (which assume independent and identically distributed errors) do not apply.

Because the physically-based deterministic models used for weather and climate prediction require “gridded” initial conditions, scientists have long been interested in methods to estimate atmospheric variables on regular grids. This problem is exactly the spatial prediction problem addressed by “kriging”. In fact, L.S. Gandin developed a complete theory of spatial best linear unbiased prediction in the context of the spatial prediction (gridding) problem in meteorology, which he called optimum interpolation (at the same time that Matheron was developing the foundations of kriging). The meteorological spatial prediction problem has some unique features that make its application different than in traditional geostatistics. First, the data come in space and time, effectively providing replications (although correlated in time) from which to deduce potentially nonstationary covariance structures. Furthermore, given that the physics governing the climate processes are relatively well-known (at some scales), spatial prediction must adhere to certain dynamical constraints. Thus, a substantial

component of the spatial prediction problem in this discipline is devoted to including appropriate dynamic constraints. The procedure for generating dynamically consistent initial conditions for deterministic geophysical models is known as data assimilation. Current work in the area focuses on performing the spatio-temporal prediction in three spatial dimensions and time, so-called four-dimensional data assimilation. The most promising methods use variational approaches or Kalman filters. Although these efforts have traditionally been tied to operational numerical weather prediction, they are also used to develop large dynamically consistent “data” sets for climate analysis and prediction.

A key challenge in climate research is the extraction of information from the huge spatio-temporal datasets generated by the data assimilation process, deterministic model output, and raw observation networks. These data sets comprise observations of extremely complicated multivariate processes. Thus, methods of analysis must be able to account for multiscale dynamical variability across different dynamical variables in space and time, account for various sources of error, and provide efficient dimension reduction. There are several key methodologies that have proven to be essential to this task. Namely, empirical orthogonal function (EOF) analysis, principal oscillation pattern (POP) analysis, spatio-temporal canonical correlation analysis (CCA), and space-time spectral analysis. In the remainder of this article, these methods are described in detail. The interested reader should note that there are several excellent reference books that have been written on statistics in the climatological sciences. A partial list is included in the Bibliography.

2 Descriptive Statistical Methods

2.1 Empirical Orthogonal Function (EOF) Analysis

EOF analysis is the geophysicist’s manifestation of the classic eigenvalue/eigenvector decomposition of a correlation (or covariance) matrix. In its discrete formulation, EOF analysis is simply Principal Component Analysis (PCA), while in the continuous framework, it is simply a Karhunen-Loève (K-L) expansion. Depending on the application, EOFs are usually used (1) in a diagnostic mode to find principal (in terms of explanation of variance) spatial structures, along with the corresponding time variation of these structures, and (2) to reduce the dimension (spatially) in large geophysical data sets while simultaneously reducing noise.

One finds in the meteorological literature, extensive use of EOFs since their introduction by Lorenz in the mid 1950s. For example, they have been used for describing climate, comparing simulations of general circulation models, de-

veloping regression forecast techniques, weather classification, map typing, the interpretation of geophysical fields, and the simulation of random fields, particularly non-homogeneous processes. In addition, as in the psychometric literature for PCAs, orthogonal and oblique rotation of EOFs often aids in the interpretation of meteorological data. Furthermore, because EOFs have difficulty resolving traveling wave disturbances, complex EOF analysis was introduced in the early 1970s and has proven to be very useful in applications to climatological analysis.

2.1.1 Continuous K-L Formulation

Consider a continuous spatial process measured at discrete time intervals. The goal is to find an optimal and separable orthogonal decomposition of a spatio-temporal process $Z(\mathbf{s}; t)$, where \mathbf{s} denotes a spatial location in some spatial domain in Euclidean space, and $t \in \{1, 2, \dots, T\}$ denotes some time. That is, consider

$$Z(\mathbf{s}; t) = \sum_{k=1}^{\infty} a_k(t) \phi_k(\mathbf{s})$$

such that $\text{var}[a_1(t)] > \text{var}[a_2(t)] > \dots$, and $\text{cov}[a_i(t), a_k(t)] = 0$ for all $i \neq k$. A well-known solution to this problem is obtained through a Karhunen-Loève (K-L) expansion. Suppose, $E[Z(\mathbf{s}; t)] = 0$, and define the covariance function as $E[Z(\mathbf{s}; t)Z(\mathbf{r}; t)] \equiv c_0^Z(\mathbf{s}, \mathbf{r})$, which need not be stationary in space, but is assumed to be invariant in time. The K-L expansion then allows the covariance function to be decomposed as follows:

$$c_0^Z(\mathbf{s}, \mathbf{r}) = \sum_{k=1}^{\infty} \lambda_k \phi_k(\mathbf{s}) \phi_k(\mathbf{r}),$$

where $\{\phi_k(\cdot) : k = 1, \dots, \infty\}$ are the eigenfunctions and $\{\lambda_k : k = 1, \dots, \infty\}$ are the associated eigenvalues of the Fredholm integral equation

$$\int_D c_0^Z(\mathbf{s}, \mathbf{r}) \phi_k(\mathbf{s}) d\mathbf{s} = \lambda_k \phi_k(\mathbf{r}), \quad (1)$$

where

$$\int_D \phi_k(\mathbf{s}) \phi_l(\mathbf{s}) d\mathbf{s} = \begin{cases} 1 & \text{for } k = l \\ 0 & \text{otherwise.} \end{cases}$$

Assuming completeness of the eigenfunctions, one can expand $Z(\mathbf{s}; t)$ according to

$$Z(\mathbf{s}; t) = \sum_{k=1}^{\infty} a_k(t) \phi_k(\mathbf{s}), \quad (2)$$

where $\{\phi_k(\mathbf{s}) : \mathbf{s} \in D\}$ is known as the k -th EOF and the associated time series $a_k(t)$ is the k -th principal component time series, or “amplitude” time series. This time series is derived from the projection of the Z process onto the EOF basis,

$$a_k(t) = \int_D Z(\mathbf{s}; t) \phi_k(\mathbf{s}) ds.$$

It is easy to verify that these time series are uncorrelated, with variance equal to the corresponding eigenvalues; that is, $E[a_i(t)a_k(t)] = \delta_{ik}\lambda_k$, where δ_{ik} equals one when $i = k$, and zero otherwise.

If the expansion (2) is truncated at K , yielding

$$Z_K(\mathbf{s}; t) \equiv \sum_{k=1}^K a_k(t) \phi_k(\mathbf{s}),$$

then it can be shown that the finite EOF decomposition minimizes the variance of the truncation error, $E\{[Z(\mathbf{s}; t) - Z_K(\mathbf{s}; t)]^2\}$, and is thus optimal in this regard when compared to all other basis sets.

Since data are always discrete, in practice one must solve numerically the Fredholm integral equation (1) to obtain the EOF basis functions. For example, numerical quadrature approaches for discretizing the integral equation succeed in that they give estimates for the eigenfunctions and eigenvalues that are weighted according to the spatial distribution of the data locations, but only for the eigenfunctions at locations $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ for which there are data. Alternatively, one can discretize the K-L integral equation and interpolate the eigenfunctions to locations where data are not available.

2.1.2 Discrete EOF Analysis

Although the continuous K-L representation of EOFs is the most realistic from a physical point-of-view, it is only rarely considered in applications. This is due simply to the discrete nature of data observations and the added difficulty of solving the K-L integral equation. Consider a discrete EOF analysis by using the PCA formulation as given in standard multivariate statistics textbooks, but according to the spatio-temporal notation introduced here. In that case, let $\mathbf{Z}(t) \equiv (Z(\mathbf{s}_1; t), \dots, Z(\mathbf{s}_n; t))'$ and define the k -th EOF ($k = 1, \dots, n$) to be $\boldsymbol{\psi}_k \equiv (\psi_k(\mathbf{s}_1), \dots, \psi_k(\mathbf{s}_n))'$, where $\boldsymbol{\psi}_k$ is the vector in the linear combination $a_k(t) = \boldsymbol{\psi}_k' \mathbf{Z}(t)$. Furthermore, $\boldsymbol{\psi}_1$ is the vector that allows $\text{var}[a_1(t)]$ to be maximized subject to the constraint $\boldsymbol{\psi}_1' \boldsymbol{\psi}_1 = 1$. Then $\boldsymbol{\psi}_2$ is the vector that maximizes $\text{var}[a_2(t)]$ subject to the constraint $\boldsymbol{\psi}_2' \boldsymbol{\psi}_2 = 1$ and $\text{cov}[a_1(t), a_2(t)] = 0$. Thus, $\boldsymbol{\psi}_k$ is the vector that maximizes $\text{var}[a_k(t)]$ subject to the orthogonality constraint

$\boldsymbol{\psi}'_k \boldsymbol{\psi}_k = 1$ and $\text{cov}[a_k(t), a_j(t)] = 0$ for all $k \neq j$. This is equivalent to solving the eigensystem $\mathbf{C}_0^Z \boldsymbol{\Psi} = \boldsymbol{\Psi} \boldsymbol{\Lambda}$, where $\mathbf{C}_0^Z \equiv \text{E}[\mathbf{Z}(t)\mathbf{Z}(t)']$, $\boldsymbol{\Psi} = (\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_n)$, $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$, and $\text{var}[a_i(t)] = \lambda_i$, $i = 1, \dots, n$. The solution is obtained by a symmetric decomposition of \mathbf{C}_0^Z , given by $\mathbf{C}_0^Z = \boldsymbol{\Psi} \boldsymbol{\Lambda} \boldsymbol{\Psi}'$.

It is straightforward to show that if a discretization of the K-L integral equation assumes equal areas of influence for each observation location, then such a discretization is equivalent to the PCA formulation of EOFs. Conversely, an EOF decomposition of irregularly spaced data without consideration of the relative area associated with each observation location leads to improper weighting of the significance of each element of the covariance matrix \mathbf{C}_0^Z . This can give erroneous results in the EOF analysis. The distinction between EOFs on a regular grid and on an irregular grid is the source of many incorrect applications of the technique in the literature.

2.1.3 Estimation of EOFs

Since the EOF analysis depends on the decomposition of a covariance matrix, we must estimate this matrix in practice. The traditional approach is based on the method of moments (MOM) estimation procedure. For example, in the discrete case with equally spaced observations, one needs an estimate of \mathbf{C}_0^Z . The MOM estimator for an element of \mathbf{C}_0^Z , is given by

$$\hat{c}_0^Z(\mathbf{s}_i, \mathbf{s}_j) \equiv (1/T) \sum_{t=1}^T [Z(\mathbf{s}_i; t) - \hat{\mu}_Z(\mathbf{s}_i; t)][Z(\mathbf{s}_j; t) - \hat{\mu}_Z(\mathbf{s}_j; t)],$$

where $\hat{\mu}_Z(\mathbf{s}_i; t)$ is an estimate of the mean of $Z(\mathbf{s}_i; t)$, for $i = 1, \dots, n$. This mean correction must be included since data usually show a nonzero mean. Typically, investigators use the estimated “time mean”,

$$\hat{\mu}_Z(\mathbf{s}_i) \equiv (1/T) \sum_{t=1}^T Z(\mathbf{s}_i; t),$$

although other estimators can be used, depending on the application.

Given an estimate $\hat{\mathbf{C}}_0^Z$ of \mathbf{C}_0^Z that is symmetric and non-negative definite (so that all eigenvalues are greater than or equal to zero), an estimate of its eigenvectors and eigenvalues can be obtained through the diagonalization $\hat{\mathbf{C}}_0^Z = \hat{\boldsymbol{\Psi}} \hat{\boldsymbol{\Lambda}} \hat{\boldsymbol{\Psi}'}$. Approximate formulas for the bias and variance of the standard eigenvalue estimator can be obtained. In general, the sample eigenvalue $\hat{\lambda}_k$ is a biased estimator of λ_k and the bias is positive for the larger λ_k 's and negative for the smaller λ_k 's. Unbiased estimators can be constructed, but the decrease in bias is

accompanied by an increase in the variance of the estimator. Furthermore, the sampling error associated with the estimated EOFs leads to numerical instability in the eigenvectors. This has led to sampling-based selection strategies for the truncation level, K .

2.1.4 Complex EOF Analysis

Consider a spatio-temporal process consisting of a sinusoid in one spatial dimension that is invariant in time, $Z(s; t) = B \sin(ls)$, where s is some location in one-dimensional space, t is a time index, B is an amplitude coefficient, and l is the spatial wave number, which is related to the wavelength L such that $l = 2\pi/L$. Now, consider the same sinusoid but allow it to have a temporal phase component (i.e., it can be considered as a wave in space which propagates in time):

$$Z(s; t) = B \sin(ls + \omega t) = B \cos(\omega t) \sin(ls) - B \sin(\omega t) \cos(ls), \quad (3)$$

where ω is the temporal frequency. In order to characterize the phase propagation of such a sinusoid, information is needed regarding the coefficients of the *two* components, $\sin(ls)$ and $\cos(ls)$, which are a quarter of a cycle out of phase. In time series analysis, this is analogous to the need for both the quadrature and co-spectrum between two time series in order to determine their spectral coherence and phase relationships.

One advantage of the EOF approach described previously is its ability to compress the complicated variability of the original data set onto a relatively small set of eigenvectors. Unfortunately, such an EOF analysis only detects spatial structures that do not change position in time. To extend the EOF analysis to the study of spatial structures that can propagate in time, one can perform a complex principal component analysis in the frequency domain. The technique involves the computation of complex eigenvectors from cross-spectral matrices. The limitation of this technique is that it only gives the decomposition for individual (i.e., very narrow) frequency bands. Consequently, if the power of a phenomenon is spread over a wide frequency band (as is generally the case with physical phenomena), then several EOF spatial maps (one for each spectral estimate) are needed to evaluate the phenomenon. This complicates the physical interpretation.

Complex empirical orthogonal function (CEOF) analysis in the time domain was developed as an alternative to the frequency-domain approach described above. This method differs from the frequency-domain approach in that Hilbert transforms (see below) are used to shift the time series of the data at each location by a quarter cycle. Analogous to (3), the original data and its Hilbert transform

allow the examination of propagating disturbances.

Consider $\{Z(\mathbf{s}_j; t) ; j = 1, \dots, n\}$ as described previously. Under certain regularity conditions, $Z(\mathbf{s}_j; t)$ has a Fourier representation of the form

$$Z(\mathbf{s}_j; t) = \sum_{\omega} \alpha_j(\omega) \cos(\omega t) + \beta_j(\omega) \sin(\omega t),$$

where $\alpha_j(\omega)$ and $\beta_j(\omega)$ are the Fourier coefficients, and ω is the frequency ($-\pi \leq \omega \leq \pi$). Since the description of propagating features requires phase information, it is convenient to use the complex representation:

$$Z^f(\mathbf{s}_j; t) = \sum_{\omega} g_j(\omega) e^{-i\omega t}, \quad (4)$$

where $g_j(\omega) = \alpha_j(\omega) + i\beta_j(\omega)$. Expanding (4) gives $Z^f(\mathbf{s}_j; t) = Z(\mathbf{s}_j; t) + i\tilde{Z}(\mathbf{s}_j; t)$, where

$$Z(\mathbf{s}_j; t) = \alpha_j(\omega) \cos(\omega t) + \beta_j(\omega) \sin(\omega t),$$

and,

$$\tilde{Z}(\mathbf{s}_j; t) = \beta_j(\omega) \cos(\omega t) - \alpha_j(\omega) \sin(\omega t).$$

The real part $Z(\mathbf{s}_j; t)$ is the original process and the imaginary part $\tilde{Z}(\mathbf{s}_j; t)$ is the Hilbert transform of the original process, which is just the original process with its phase shifted in time by $\pi/2$.

Now, the covariance matrix of $Z^f(\mathbf{s}_j; t)$ can be written as:

$$\mathbf{C}_0^{Z^f} = [c_0^{Z^f}(\mathbf{s}_j, \mathbf{s}_k)]_{j,k=1,\dots,n}$$

where $c_0^{Z^f}(\mathbf{s}_j, \mathbf{s}_k) \equiv \mathbb{E}[Z^f(\mathbf{s}_j; t)^* Z^f(\mathbf{s}_k; t)]$ (assuming zero mean), and where $*$ denotes the complex conjugate. Note that $\mathbf{C}_0^{Z^f}$ is essentially the cross-spectral matrix averaged over all frequencies ($-\pi \leq \omega \leq \pi$), and thus leads to an average depiction of the propagating disturbances present in the data. If interest concerns phenomena occurring over a certain spectral frequency range of ω , then the original process $Z(\cdot; t)$ and its Hilbert transform $\tilde{Z}(\cdot; t)$ can be filtered accordingly before the CEOF analysis.

Since $\mathbf{C}_0^{Z^f}$ is Hermitian, it possesses real eigenvalues $\{\lambda_k\}$ and complex eigenvectors, $\boldsymbol{\gamma}_k \equiv (\gamma_k(\mathbf{s}_1), \dots, \gamma_k(\mathbf{s}_n))'$ where $k = 1, \dots, n$. The EOF representation of $Z^f(\cdot; t)$, which optimally accounts for the variance of $Z(\cdot; t)$ in the frequency band of interest, is:

$$Z^f(\mathbf{s}_i; t) = \sum_{k=1}^n a_k(t) \boldsymbol{\gamma}_k^*(\mathbf{s}_i),$$

where the complex time-dependent principal components are given by:

$$a_k(t) = \sum_{i=1}^n Z^f(\mathbf{s}_i; t) \gamma_k(\mathbf{s}_i).$$

Four measures are generally used to examine the structure of the CEOFs.

- *Spatial Phase Function.* The spatial phase function is given by:

$$\theta_k(\mathbf{s}_i) = \arctan \left[\frac{\text{Im}(\gamma_k(\mathbf{s}_i))}{\text{Re}(\gamma_k(\mathbf{s}_i))} \right].$$

This function can take any value between $-\pi$ and π . In the case of the simple sinusoid with temporal phase (3), this corresponds to ls . In that case the spatial phase will go through one complete cycle (2π) over the distance $2\pi/l$. Note that for data fields that include many different scales of variability, the spatial phase plot can be very difficult to interpret. Pre-filtering generally improves interpretability.

- *Spatial Amplitude Function.* The spatial amplitude function is given by:

$$S_k(\mathbf{s}_i) = [\gamma_k(\mathbf{s}_i) \gamma_k^*(\mathbf{s}_i)]^{1/2}.$$

This function is interpreted in the same way as the eigenfunctions in traditional EOF analysis.

- *Temporal Phase Function.* The temporal phase function is given by:

$$\xi_k(t) = \arctan \left[\frac{\text{Im}(a_k(t))}{\text{Re}(a_k(t))} \right].$$

Consider the simple sinusoid example in (3). For a fixed frequency ω_0 , this temporal phase function would give $\omega_0 t$ (i.e., a linear relationship in time). In practice, this provides information as to the frequency of the dominant component of a particular eigenvector at a given time.

- *Temporal Amplitude Function.* The temporal amplitude function is given by:

$$R_k(t) = [a_k(t) a_k^*(t)]^{1/2}.$$

This function corresponds to the amplitude time series as given in traditional EOF analysis.

2.1.5 Multivariate EOF Analysis

Often, one may be interested in the simultaneous analysis of two or more processes. Consider two fields observed over time at the same spatial locations; that is, consider $Z(\mathbf{s}_i; t)$ and $X(\mathbf{s}_i; t)$, where $i = 1, \dots, n$; $t = 1, \dots, T$. Then, write $\mathbf{W}(t) \equiv [\mathbf{Z}(t)' \mathbf{X}(t)']'$, where $\mathbf{X}(t) \equiv (X(\mathbf{s}_1; t), \dots, X(\mathbf{s}_n; t))'$, $\mathbf{Z}(t)$ is defined as before, and the covariance matrix of $\mathbf{W}(t)$ is given by \mathbf{C}_0^W . This matrix includes off-diagonal submatrices that represent the covariance between $\mathbf{Z}(t)$ and $\mathbf{X}(t)$. One can then obtain the EOF solution in the conventional manner by diagonalizing the \mathbf{C}_0^W matrix; that is, $\mathbf{C}_0^W = \mathbf{\Psi}_W \mathbf{\Lambda}_W \mathbf{\Psi}_W'$, where the columns of $\mathbf{\Psi}_W$ are the eigenvectors (i.e., EOFs) and $\mathbf{\Lambda}_W$ is a diagonal matrix containing the eigenvalues of \mathbf{C}_0^W . Then, the first n elements of the k -th eigenvector correspond to the portion of the k -th EOF for the Z process, and the last n elements correspond to the portion representative of the k -th EOF of the X process. Theoretically, there is no limit to the number of processes that can be considered simultaneously. However, there is a practical limitation to this procedure since the covariance matrix can easily become very large if the number of observation locations or variables increases. Comparisons of this approach to other multivariate methods such as CCA suggests that in some cases the multivariate EOF approach has large biases and does not perform well in small signal-to-noise ratio situations.

2.1.6 Extended EOF Analysis

Extended EOFs are simply multivariate EOFs in which the additional variables are lagged versions of the same process. For example, we could let $\mathbf{W}(t) = [\mathbf{Z}(t)' \mathbf{Z}(t-1)']'$. In this case, if temporal invariance is assumed, then the diagonal sub-matrices of \mathbf{C}_0^W are equivalent, and the off-diagonal submatrices are just the lag-one correlation matrices $\mathbf{C}_1^Z \equiv \text{cov}[\mathbf{Z}(t), \mathbf{Z}(t-1)]$. In this way, one can examine the propagation of EOF spatial patterns in time by noting that the first n eigenvalues of a particular eigenvector correspond to the time zero representation of that EOF, and the second n eigenvalues correspond to the lag one representation of the same EOF. This approach is closely linked with time-lagged CCA and the minimum/maximum autocorrelation factor (MAF) method in statistics.

2.2 Principal Oscillation Patterns (POP) Analysis

In essence, Principal Oscillation Pattern (POP) analysis assumes that the (spatial) data field has a temporal autoregressive structure of order one, AR(1). A chief difference between POPs and other spatio-temporal decompositions is that the eigenvectors (i.e., the spatial patterns) are not orthonormal. In addition,

the eigenvectors can be characterized as the empirical “normal modes” of the (estimated) system matrix of the fitted AR(1) stochastic process.

2.2.1 Formulation of POPs

Hypothesize the following vector AR(1) model for the dynamical process $\mathbf{Z}(t)$:

$$\mathbf{Z}(t+1) = \mathbf{B}\mathbf{Z}(t) + \boldsymbol{\eta}(t), \quad (5)$$

where \mathbf{B} is an $n \times n$ real matrix (possibly non-symmetric), $\boldsymbol{\eta}(t)$ is an $n \times 1$ additive error vector (often assumed to be Gaussian) such that $E[\boldsymbol{\eta}(t)\mathbf{Z}(t)'] = \mathbf{0}$, $E[\boldsymbol{\eta}(t)] = \mathbf{0}$, and $E[\boldsymbol{\eta}(t)\boldsymbol{\eta}(\tau)'] \equiv \mathbf{C}^\eta$ when $t = \tau$, and $\mathbf{0}$ otherwise. From well-known results in multivariate time series, $\mathbf{B} = \mathbf{C}_1^Z(\mathbf{C}_0^Z)^{-1}$, where $\mathbf{C}_1^Z \equiv \text{cov}[\mathbf{Z}(t+1), \mathbf{Z}(t)]$, $\mathbf{C}_0^Z \equiv \text{var}[\mathbf{Z}(t)]$, and it has been assumed that \mathbf{C}_0^Z is non-singular.

From well-known results in matrix algebra, a non-symmetric matrix $\mathbf{A}_{n \times n}$ has characteristic equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$, with n roots, some of which may be complex even if \mathbf{A} is real. Corresponding to a latent root λ_i there are two vectors $\mathbf{p}_i, \mathbf{q}_i$ called the “right” and “left” singular vectors, such that $\mathbf{A}\mathbf{p}_i = \lambda_i\mathbf{p}_i$ and $\mathbf{A}'\mathbf{q}_i = \lambda_i\mathbf{q}_i$ for $i = 1, \dots, n$. It can be shown that (i) $\{\mathbf{p}_i\}$ are linearly independent and so are $\{\mathbf{q}_i\}$, and (ii) $\mathbf{p}_i^*\mathbf{q}_j = 0$ for $i \neq j$, where $*$ denotes the Hermitian transpose (e.g., $\mathbf{a}^*\mathbf{b} = \sum_i a_i b_i^*$). In addition, if we let $\mathbf{p}_i^*\mathbf{q}_i \equiv d_i$ and $\mathbf{D} \equiv \text{diag}(d_i): i = 1, \dots, n$, then \mathbf{A} has the spectral decomposition (assuming that the latent roots are distinct),

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{D}^{-1}\mathbf{Q}' = \sum_{i=1}^n \frac{\lambda_i}{d_i} \mathbf{p}_i \mathbf{q}_i^*. \quad (6)$$

Using (6), decompose the AR(1) model matrix \mathbf{B} in (5) as

$$\mathbf{B} = \sum_{i=1}^n \frac{\lambda_i}{d_i} \boldsymbol{\phi}_i \boldsymbol{\psi}_i^*,$$

where $d_i \equiv \boldsymbol{\psi}_i^* \boldsymbol{\phi}_i$, $\boldsymbol{\phi}_i$ is the right singular vector of \mathbf{B} , and $\boldsymbol{\psi}_i$ is the left singular vector of \mathbf{B} , corresponding to the latent root $\lambda_i: i = 1, \dots, n$. Because

$$\sum_{i=1}^n \frac{\boldsymbol{\phi}_i \boldsymbol{\psi}_i^*}{d_i} = \boldsymbol{\Phi}\mathbf{D}^{-1}\boldsymbol{\Psi}^* = \boldsymbol{\Phi}(\boldsymbol{\Psi}^*\boldsymbol{\Phi})^{-1}\boldsymbol{\Psi}^* = \mathbf{I}_{n \times n},$$

it follows that

$$\mathbf{Z}(t) = \sum_{i=1}^n \frac{\boldsymbol{\phi}_i \boldsymbol{\psi}_i^*}{d_i} \mathbf{Z}(t) = \sum_{i=1}^n a_i(t) \boldsymbol{\phi}_i, \quad (7)$$

where

$$a_i(t) \equiv \frac{\boldsymbol{\psi}_i^* \mathbf{Z}(t)}{d_i}. \quad (8)$$

The essence of POP analysis is the decomposition represented by (7) and (8). The vectors $\{\boldsymbol{\phi}_i\}$ are called *principal oscillation patterns* and, although they constitute a linear basis, they are not orthonormal. The time series $a_i(t)$ are known as POP coefficients. Although the $\boldsymbol{\phi}_i$ are not orthogonal with themselves, they are orthogonal with the normalized adjoint patterns $\boldsymbol{\psi}_i^*/d_i$.

2.2.2 Physical Implication of POPs

To gain insight into the physical meaning of the POPs, consider an idealized discrete linear system (i.e., with no error term),

$$\mathbf{Z}(t+1) = \mathbf{B}\mathbf{Z}(t). \quad (9)$$

The eigenvectors $\boldsymbol{\phi}_i$; $i = 1, \dots, n$ of the \mathbf{B} matrix are referred to as the system *normal modes*. Because \mathbf{B} is not in general symmetric, some or all of its eigenvalues and eigenvectors are complex. Furthermore, because \mathbf{B} is real, the complex conjugates λ_i^* and $\boldsymbol{\phi}_i^*$ also satisfy the eigen-equation $\mathbf{B}\boldsymbol{\phi}_i^* = \lambda_i^* \boldsymbol{\phi}_i^*$. Now, multiplying (9) by $\boldsymbol{\psi}_j^*/d_j$ gives

$$a_j(t+1) = \frac{\boldsymbol{\psi}_j^*}{d_j} \left[\sum_{i=1}^n \frac{\lambda_i}{d_i} \boldsymbol{\phi}_i \boldsymbol{\psi}_i^* \right] \mathbf{Z}(t) = \frac{\boldsymbol{\psi}_j^*}{d_j} \boldsymbol{\phi}_j \lambda_j a_j(t) = \lambda_j a_j(t),$$

which is a first-order homogeneous difference equation with solution $a_j(t) = (\lambda_j)^t$ (assuming $a_j(0) = 1$). If λ_j is complex and $i \equiv \sqrt{-1}$, then $\lambda_j \equiv \lambda_j^R + i\lambda_j^I$, which can be written in polar form as $\lambda_j^R = \gamma_j \cos(\omega_j)$ and $\lambda_j^I = \gamma_j \sin(\omega_j)$. Then, $\lambda_j = \gamma_j e^{i\omega_j}$, where $\gamma_j = \{(\lambda_j^R)^2 + (\lambda_j^I)^2\}^{1/2}$. Thus, $a_j(t) = \gamma_j^t e^{i\omega_j t}$ which, under stationarity conditions ($|\lambda_j| \leq 1$; $j = 1, \dots, n$), shows that $a_j(t)$ evolves as a damped spiral in the complex plane with a characteristic damping rate γ_j and frequency ω_j .

Decompose the eigenvector (i.e., normal mode) $\boldsymbol{\phi}_j$ as the sum of a real and an imaginary term $\boldsymbol{\phi}_j = \boldsymbol{\phi}_j^R + i\boldsymbol{\phi}_j^I$. Note that for \mathbf{B} real, the normal modes occur in complex conjugate pairs (if they are complex at all), and the general evolution of a damped normal mode (i.e., $\gamma_j \leq 1$) can be described in a two-dimensional subspace spanned by $\boldsymbol{\phi}_j^R$ and $\boldsymbol{\phi}_j^I$. That is, the evolution of a damped mode occurs in a succession

$$\dots \rightarrow \boldsymbol{\phi}_j^R \rightarrow -\boldsymbol{\phi}_j^I \rightarrow -\boldsymbol{\phi}_j^R \rightarrow \boldsymbol{\phi}_j^I \rightarrow \boldsymbol{\phi}_j^R \rightarrow \dots \quad (10)$$

with a period of $2\pi/\omega_j$, each stage in (10) occurring a quarter of a cycle apart. The time τ_j needed to reduce an initial *amplitude* $a_j(0)$ to $a_j(0)/\exp(1)$ is referred to as the *e-folding time* and is given by $\tau_j \equiv -1/\ln(\gamma_j)$.

2.2.3 Estimation of POPs

The previous section emphasized the physical motivation behind the POP analysis. If the system were *deterministic*, then the normal mode approach would be sufficient. However, the AR(1) representation is *stochastic* and, as such, the effect of the error process must be considered.

In order to perform the POP analysis in practice, the system matrix \mathbf{B} must be estimated. A MOM estimator for \mathbf{B} is $\hat{\mathbf{B}} = \hat{\mathbf{C}}_1^Z [\hat{\mathbf{C}}_0^Z]^{-1}$, where $\hat{\mathbf{C}}_0^Z$ is the MOM estimator described previously. Similarly, the (i, j) -th element of \mathbf{C}_1^Z is given by, $c_1^Z(\mathbf{s}_i, \mathbf{s}_j) \equiv \text{E}[Z(\mathbf{s}_i; t)Z(\mathbf{s}_j; t - 1)]$, so that the MOM estimator is

$$\hat{c}_1^Z(\mathbf{s}_i, \mathbf{s}_j) \equiv \frac{1}{T-1} \sum_{t=2}^T [Z(\mathbf{s}_i; t) - \hat{\mu}_z(\mathbf{s}_i; t)][Z(\mathbf{s}_j; t-1) - \hat{\mu}_z(\mathbf{s}_j; t-1)].$$

The decomposition of $\hat{\mathbf{B}}$ then gives estimated eigenvectors $\hat{\phi}_j$, adjoints $\hat{\psi}_j$, and eigenvalues $\hat{\lambda}_j$, $j = 1, \dots, n$. The estimated eigenvectors are sometimes referred to as *empirical* normal modes, analogous to the deterministic decomposition. It is then assumed, sometimes erroneously, that these empirical normal modes behave as one would expect the deterministic normal modes to behave. For example, for damped empirical modes (i.e., $\hat{\gamma}_j \leq 1$, where $\hat{\gamma}_j = |\hat{\lambda}_j|$) one expects the succession (10). However, in the presence of error, this relationship may not hold. To check if this relationship is valid in practice, a cross-spectral analysis is often performed between the estimates of the real component $a_j^R(t)$ and the imaginary component $a_j^I(t)$ of $a_j(t)$ which, according to the deterministic analysis, should vary coherently with a frequency ω_j and phase lag $\pi/2$, $a_j^R(t)$ lagging $a_j^I(t)$.

2.2.4 Diagnostic Applications of POPs

In a diagnostic mode, POPs are used to examine the oscillation properties and spatial structure of dynamical processes in the atmosphere. In this case, one looks at the estimated frequencies $\{\hat{\omega}_j\}$, amplitudes $\{\hat{\gamma}_j\}$, and e-folding times $\{\hat{\tau}_j\}$, as well as the amplitude time series $\{\hat{a}_j(t)\}$ and eigenvectors $\{\hat{\phi}_j\}$. These quantities give insight into the physical meaning behind the empirical normal modes. Often, one first filters the data (usually in time) to focus on a particular atmospheric phenomenon. Essentially, this is an *ad hoc* method for removing the error, so that the deterministic interpretation is more tenable. Clearly, there is

no guarantee that such a noise-reduction scheme is optimal (almost certainly, it is not), but the issue has not been addressed in the literature. After filtering, it is hoped that the spatial patterns of the empirical normal modes can illustrate the spatial structure of the phenomenon of interest as it evolves in time.

2.2.5 Prognostic Applications of POPs

Since POPs have an inherent AR(1) dynamical structure, they are ideally suited for prognostic applications. Analogous to the deterministic case, it is easy to show from (5) that $a_j(t+1) = \lambda_j a_j(t) + \tilde{\eta}_j(t)$, where $\tilde{\eta}_j(t) \equiv \boldsymbol{\psi}_j^* \boldsymbol{\eta}(t)/d_j$. The presence of the noise term clearly prevents the use of the deterministic normal mode results for prediction. However, this noise $\tilde{\eta}_j(t)$ is typically ignored in practice. The justification is that usually only one empirical normal mode is assumed to be of physical importance, so after prefiltering in favor of this mode, the noise is assumed to be negligible. This is a tenuous assumption. In spite of these reservations, predictions can be obtained from $\hat{a}_j(t+1) = \hat{\gamma}_j e^{i\omega_j} \hat{a}_j(t)$, which then gives $\hat{\mathbf{Z}}_j(t+1) \equiv \hat{a}_j(t+1) \boldsymbol{\phi}_j$, and hence $\hat{\mathbf{Z}}(t+1) = \sum_{j=1}^n \hat{\mathbf{Z}}_j(t+1)$, or perhaps a truncated version. In the deterministic case, if we knew the location in the complex state space of the system at any given time, we could predict perfectly into the future. Clearly, the presence of noise limits the skill of any such approach. However, it is argued that even in the presence of “unpredictable noise”, such a scheme should be useful for short time leads.

We note that the POP formulation and prediction methodology with a stationary model inherently assumes a decay in amplitude (since stationarity implies $\gamma_j \leq 1$; $j = 1, \dots, n$). Thus, it is common to “respecify” the estimate of γ_j , by setting it equal to one. In this case, the amplitude does not change with time (i.e., a persistence forecast of amplitude is assumed). Then, the frequency ω_j takes on added importance since the choice of initial phase becomes critical. It is usually the case that $\mathbf{Z}(T)$ (i.e., the latest observations) are noisy, so that $a_j(T)$ is too noisy to use in the prediction. Thus, some form of noise reduction is applied. Typically, this entails projecting the data onto a limited set of the first K EOFs (to smooth the data in space) and to apply a time filter.

As mentioned above, the linear stationary nature of the POP methodology forces all oscillatory solutions to decay. In a diagnostic analysis this does not present a problem (and, actually, the decay rate can be useful information). However, it would seem to be quite inappropriate in the forecasting framework, particularly for the atmosphere, where most phenomena have amplitudes that are growing at some point in their evolution. Some authors make a strong case for using a large set of empirical normal modes in the forecast. In that case, constructive interference between the various empirical modes allows for the growth of

certain multi-mode phenomena. Physically, this approach has much more appeal than the single mode approach.

2.2.6 POPs in Continuous Time

From a physical viewpoint, the time domain in the POP analysis should be considered as continuous. In that case, one must solve the appropriate Fokker-Planck equation to get a probabilistic solution to the stochastic differential equation describing the temporal evolution of the dynamical process of interest.

2.2.7 Complex POPs

Complex POP (CPOP) analysis has been considered as an extension of conventional POP analysis. While POP analysis is able to model *traveling oscillations*, they are unable to model *standing oscillations*. Thus, CPOP analysis is a natural extension to POP analysis, in many ways analogous to the relationship between EOFs and CEOFs. That is, CEOFs are able to detect traveling disturbances which cannot be detected by EOFs, and CPOPs can detect standing oscillations, which cannot be detected by POPs.

2.2.8 Cyclostationary POPs

Traditional time series analysis techniques (including the vector autoregression case that is analogous to POPs) require an assumption of second-order stationarity (i.e., constant mean with autocorrelation depending only on time lag). This assumption clearly breaks down when the physical process under consideration has known cycles (i.e., solar influenced annual and semiannual cycles in atmospheric processes). In that case the mean and variance (at least) are also periodic. Traditionally, investigators remove these cycles, hoping that they then can satisfy the stationarity assumption (which typically, they cannot, at least with regard to the variance). However, from a statistical perspective, it makes sense to use the redundant information contained in the periodically correlated statistics optimally, rather than to remove it. Some authors have considered using periodically correlated statistics in the POP formulation.

2.3 Space-Time Canonical Correlation Analysis (CCA)

Canonical Correlation Analysis (CCA) is a long-standing multivariate statistical technique that finds linear combinations of two sets of random variables, whose correlations are maximal. In the atmospheric sciences, CCA has been used in

diagnostic climatological studies, in the forecast of El Niño, and the forecast of long-range temperature and precipitation.

2.3.1 Two-Field Spatial-Temporal CCA

Assume that in addition to the process $Z(\mathbf{s}; t)$ there is another related process $X(\mathbf{s}; t)$ with a possibly different spatial domain, but the same temporal domain. Further assume discrete space and time and zero means $\mathbf{E}[Z(t)] = \mathbf{0}_{n \times 1}$ and $\mathbf{E}[X(t)] = \mathbf{0}_{m \times 1}$ where $\mathbf{X}(t) \equiv (X(\mathbf{r}_1; t), \dots, X(\mathbf{r}_m; t))'$ and $\mathbf{Z}(t)$ is defined as before. Define the covariances $\mathbf{C}_0^Z \equiv \mathbf{E}[Z(t)Z(t)']_{n \times n}$, $\mathbf{C}_0^X \equiv \mathbf{E}[X(t)X(t)']_{m \times m}$, and $\mathbf{C}_0^{Z,X} \equiv \mathbf{E}[Z(t)X(t)']_{n \times m}$, which are invariant in time. One seeks linear combinations of each data field $a_k(t) = \phi_k' Z(t)$ and $b_k(t) = \psi_k' X(t)$, where $k = 1, \dots, \min\{m, n\}$.

Define the k -th *canonical correlation* as

$$r_k \equiv \text{corr}[\phi_k' Z(t), \psi_k' X(t)] = \frac{\phi_k' \mathbf{C}_0^{Z,X} \psi_k}{(\phi_k' \mathbf{C}_0^Z \phi_k)^{1/2} (\psi_k' \mathbf{C}_0^X \psi_k)^{1/2}}. \quad (11)$$

The first pair of canonical variables are defined as the set of linear combinations $a_1(t)$ and $b_1(t)$ for $\{t = 1, \dots, T\}$ that maximize the correlation (11) and have unit variance. Then, the k -th set of canonical variables are the linear combinations $a_k(t)$ and $b_k(t)$ that are uncorrelated with the previous $k - 1$ canonical pairs, have unit variance, and maximize (11).

Initially, let $k = 1$ and note that since \mathbf{C}_0^Z and \mathbf{C}_0^X are positive definite, they can be written as $\mathbf{C}_0^Z = (\mathbf{C}_0^Z)^{1/2} (\mathbf{C}_0^Z)^{1/2}$ and $\mathbf{C}_0^X = (\mathbf{C}_0^X)^{1/2} (\mathbf{C}_0^X)^{1/2}$. Then,

$$r_k^2 = \frac{[\tilde{\phi}_1' (\mathbf{C}_0^Z)^{-1/2} \mathbf{C}_0^{Z,X} (\mathbf{C}_0^X)^{-1/2} \tilde{\psi}_1]^2}{(\tilde{\phi}_1' \tilde{\phi}_1)(\tilde{\psi}_1' \tilde{\psi}_1)}, \quad (12)$$

where $\tilde{\phi}_1 \equiv (\mathbf{C}_0^Z)^{1/2} \phi_1$ and $\tilde{\psi}_1 \equiv (\mathbf{C}_0^X)^{1/2} \psi_1$ are normalized weights. The problem is now reduced to finding $\tilde{\phi}_1$ and $\tilde{\psi}_1$ that maximize (12). Note that $(\mathbf{C}_0^Z)^{-1/2} \mathbf{C}_0^{Z,X} (\mathbf{C}_0^X)^{-1/2}$ is not symmetric, which means that a singular value decomposition is needed. It can be shown that r_1^2 is the largest singular value of $(\mathbf{C}_0^Z)^{-1/2} \mathbf{C}_0^{Z,X} (\mathbf{C}_0^X)^{-1/2}$, where $\tilde{\phi}_1$ and $\tilde{\psi}_1$ are the left and right singular vectors corresponding to r_1^2 , respectively. Then, one can write $\phi_1 \equiv (\mathbf{C}_0^Z)^{-1/2} \tilde{\phi}_1$ and $\psi_1 \equiv (\mathbf{C}_0^X)^{-1/2} \tilde{\psi}_1$ and can obtain the time series of canonical variables (for $t = 1, \dots, T$), $a_1(t) = \phi_1' Z(t)$ and $b_1(t) = \psi_1' X(t)$. In general, $\tilde{\phi}_k$ and $\tilde{\psi}_k$ are the left and right singular vectors, respectively, associated with the k -th singular value r_k^2 from the singular value decomposition. Then, ϕ_k , ψ_k , $a_k(t)$, and $b_k(t)$ can be obtained analogous to the case for $k = 1$.

The correlation between the canonical variable time series and the original data can also be examined. That is,

$$r[a_k(t), \mathbf{Z}(t)] \equiv \text{corr}[a_k(t), \mathbf{Z}(t)] = \boldsymbol{\phi}'_k \mathbf{C}_0^Z [\text{diag}(\mathbf{C}_0^Z)]^{-1/2} \quad (13)$$

and

$$r[b_k(t), \mathbf{X}(t)] \equiv \text{corr}[b_k(t), \mathbf{X}(t)] = \boldsymbol{\psi}'_k \mathbf{C}_0^X [\text{diag}(\mathbf{C}_0^X)]^{-1/2}, \quad (14)$$

where $[\text{diag}(\mathbf{C}_0^Z)]^{-1/2}$ is an $n \times n$ diagonal matrix with (i, i) -th element given by $\{\text{var}[Z(\mathbf{s}_i; t)]\}^{-1/2}$, and $[\text{diag}(\mathbf{C}_0^X)]^{-1/2}$ is an $m \times m$ diagonal matrix with (i, i) -th element given by $\{\text{var}[X(\mathbf{r}_i; t)]\}^{-1/2}$.

Equation (13) and (14) are known as the k -th left and right *homogeneous correlation maps*, respectively. These are maps in the sense that the i -th value of the vector is identified with the i -th location in Euclidean space. Similarly, one can define the k -th left and right *heterogeneous correlation maps*, respectively, as

$$r[a_k(t), \mathbf{X}(t)] \equiv \text{corr}[a_k(t), \mathbf{X}(t)] = \boldsymbol{\phi}'_k \mathbf{C}_0^{Z,X} [\text{diag}(\mathbf{C}_0^X)]^{-1/2}$$

and

$$r[b_k(t), \mathbf{Z}(t)] \equiv \text{corr}[b_k(t), \mathbf{Z}(t)] = \boldsymbol{\psi}'_k \mathbf{C}_0^{X,Z} [\text{diag}(\mathbf{C}_0^Z)]^{-1/2}.$$

These maps represent how well the observations in one field can be explained by the k -th canonical variable from the other field.

2.3.2 Estimation of CCA

For the singular value decomposition, estimates of \mathbf{C}_0^Z , \mathbf{C}_0^X , and $\mathbf{C}_0^{Z,X}$ are needed. As with EOFs and POPs, estimation is performed by the method of moments:

$$\hat{\mathbf{C}}_Z \equiv \frac{1}{T} \sum_{t=1}^T \mathbf{Z}(t) \mathbf{Z}(t)', \quad \hat{\mathbf{C}}_X \equiv \frac{1}{T} \sum_{t=1}^T \mathbf{X}(t) \mathbf{X}(t)', \quad \hat{\mathbf{C}}_{Z,X} \equiv \frac{1}{T} \sum_{t=1}^T \mathbf{Z}(t) \mathbf{X}(t)',$$

where both the $Z(\cdot)$ and $X(\cdot)$ processes are assumed to have zero means. Then, the estimated singular values and singular vectors are obtained from the numerical singular value decomposition of $(\hat{\mathbf{C}}_0^Z)^{-1/2} \hat{\mathbf{C}}_0^{Z,X} (\hat{\mathbf{C}}_0^X)^{-1/2}$. The CCA from these estimated matrices is often unsatisfactory because the covariance matrix estimates are noisy when estimated with sample sizes common in the atmospheric sciences. To compensate for this, the data are often projected onto a truncated set of EOFs before applying the singular value decomposition. This is the same technique suggested for use prior to a POP analysis. As in that case, there is a benefit from the reduction of spatial dimension (clearly necessary if $\min\{n, m\} > T$, which would otherwise imply a singular covariance matrix) and computational stabil-

ity. As is always the case with truncated EOFs, there is some question about the appropriate number of EOFs to retain, and a substantial literature exists to help make this determination.

2.3.3 Time-Lagged CCA

When only one process (say $Z(\cdot)$) is considered, the CCA technique can be used to find the canonical correlation patterns between $\mathbf{Z}(t)$ and $\mathbf{Z}(t + \tau)$, for some time lag τ . Such an analysis is then useful for prognostic applications. In this case, consider the vectors ϕ_k and ψ_k such that the marginal correlation between $a_k(t) = \phi_k' \mathbf{Z}(t)$, and $b_k(t + \tau) = \psi_k' \mathbf{Z}(t + \tau)$ is maximized. To do this, obtain the k -th left and right singular vectors from the singular value decomposition applied to $(\mathbf{C}_0^Z)^{-1/2} \mathbf{C}_\tau^Z (\mathbf{C}_0^Z)^{-1/2}$, where $\mathbf{C}_\tau^Z \equiv E[\mathbf{Z}(t) \mathbf{Z}(t + \tau)']$, and where it has been assumed that $Z(\cdot)$ has zero mean. Analogous to the two-field CCA previously described, to obtain ϕ_k and ψ_k we have to weight the eigenvectors obtained from this singular value decomposition by premultiplying them with $(\mathbf{C}_0^Z)^{-1/2}$.

The time-lagged CCA approach outlined here is similar to the POP analysis and to the minimum/maximum autocorrelation factor (MAF) approach in statistics. In the MAF case, one is interested in the eigenvectors proportional to those obtained from the singular value decomposition of $(\mathbf{C}_0^Z)^{-1} \mathbf{V}$, where the matrix \mathbf{V} is the first-difference correlation matrix:

$$\begin{aligned} \mathbf{V} &\equiv E[(\mathbf{Z}(t) - \mathbf{Z}(t + 1))(\mathbf{Z}(t) - \mathbf{Z}(t + 1))'] \\ &= 2\mathbf{C}_0^Z - \mathbf{C}_1^Z - \mathbf{C}_1^{Z'} \\ &= 2\mathbf{C}_0^Z \left[\mathbf{I} - (\mathbf{C}_0^Z)^{-1} \left(\frac{\mathbf{C}_1^Z + (\mathbf{C}_1^Z)'}{2} \right) \right], \end{aligned} \quad (15)$$

where temporal invariance has been assumed in obtaining (15). Proportionality factors are used to ensure that the MAFs (i.e., $\alpha_k(t)$, where $\alpha_k(t) \equiv \phi_k' \mathbf{Z}(t)$; $k = 1, \dots, n$) have unit variance and positive correlation with time.

When considering POPs, time-lagged CCA (for $\tau = 1$), and MAFs, the singular value decomposition must be performed on the following matrices, respectively:

$$\mathbf{C}_1^Z (\mathbf{C}_0^Z)^{-1}, \quad (16)$$

$$(\mathbf{C}_0^Z)^{-1/2} \mathbf{C}_1^Z (\mathbf{C}_0^Z)^{-1/2}, \quad (17)$$

$$(\mathbf{C}_0^Z)^{-1} \left[\frac{\mathbf{C}_1^Z + (\mathbf{C}_1^Z)'}{2} \right]. \quad (18)$$

Clearly, these three matrices are similar in that they include some form of the

lag-one covariance matrix and the inverse of the lag-zero covariance matrix. Furthermore, all three are square matrices but are nonsymmetric and thus, require the singular value decomposition. It is straightforward to show (using the respective eigen-equations) that the singular values are the same for the decomposition of (16) and (17) and that the singular vectors from the CCA decomposition (17) are equivalent to those from the POP decomposition (16) scaled by the matrix $(\mathbf{C}_0^Z)^{-1/2}$. The relationship between the eigenvalues and eigenvectors of the MAF matrix (18) and the POP and CCA matrices is more complicated. However, the MAF approach does have one clear advantage. Note that although \mathbf{C}_1^Z is nonsymmetric, the matrix $\mathbf{C}_1^Z + (\mathbf{C}_1^Z)'$ is symmetric. Then, since the MAF approach is invariant to affine transformations, a symmetric matrix can be obtained in (18) by first projecting the data onto its EOF basis (so that the lag-zero covariance matrix is the identity matrix). In this case, the symmetry implies that the singular vectors associated with $\mathbf{Z}(t)$ and $\mathbf{Z}(t + \tau)$ are the same (i.e., $\boldsymbol{\psi}_k = \boldsymbol{\phi}_k; k = 1, \dots, n$), so one is able to consider the linear combinations $\boldsymbol{\phi}'_k \mathbf{Z}(t)$ and $\boldsymbol{\phi}'_k \mathbf{Z}(t + \tau)$, which is useful.

2.4 Space-Time Spectral Analysis

The spectral analysis of time series has been used extensively in climatology. Similarly, two-dimensional spectral analysis has been used in spatial problems to describe both the structure and patterns of spatial data sets. It is natural to consider a hybrid of these methodologies in climatology due to the prominence of wave-like processes in the governing dynamics. In particular, to make inference about relationships between spatial and temporal scales of variability, meteorological data are often analyzed by a variant of two-dimensional spectral analysis in which time and one spatial dimension are considered.

Assume a given spatio-temporal process $Z(x, t)$ is cyclic in longitude x (positive to the east) and limited in time $t \in [0, T]$. This process can be expanded in zonal (east-west along a latitude circle) Fourier harmonics of wavenumber k :

$$Z(x, t) = \mu(t) + \sum_{k=1}^{\infty} \{c_k(t)\cos(kx) + s_k(t)\sin(kx)\}$$

where $\mu(t)$ represents the spatial average along the latitude circle, and $c_k(t)$ and $s_k(t)$ are the time-varying longitudinal cosine and sine Fourier transform coefficients, respectively. The space-time power spectrum, P , of eastward traveling (+) and westward traveling (-) waves can be computed according to

$$P(k, \pm\omega) = \frac{1}{4} [P_\omega(c_k) + P_\omega(s_k) \pm 2Q_\omega(c_k, s_k)],$$

where ω is the normalized time frequency, P_ω is the time-power spectra, and Q_ω is the quadrature spectrum of $c_k(t)$ and $s_k(t)$. This formula is a generalization of the observation that since the cosine and sine coefficients of traveling waves are 90 degrees out of phase, the quadrature spectrum indicates the squared amplitude of traveling waves, and its sign indicates the direction of phase propagation. The space-time power spectrum of the standing oscillation, ST , can then be expressed as

$$ST(k, |\omega|) = \left\{ \frac{1}{4} [P_\omega(c_k) - P_\omega(s_k)]^2 + K_\omega^2(c_k, s_k) \right\}^{1/2},$$

where $K_\omega(c_k, s_k)$ is the co-spectrum of c_k and s_k . It is then possible to isolate the “pure” eastward and westward propagating components, $PPUR$, of a wave as

$$PPUR(k, \pm\omega) = P(k, \pm\omega) - \frac{1}{2}ST(k, |\omega|).$$

This method of partitioning space-time power spectra into standing and traveling parts is based on the following definitions and assumptions:

- standing waves are defined as consisting of eastward and westward moving components which are of equal amplitude and are coherent,
- traveling waves are defined as consisting of eastward and westward moving components which are incoherent with each other, and
- standing and traveling waves are of a different origin and are incoherent with each other.

It is important to note that the validity and applicability of these assumptions are often called into question. In addition to considering the validity of these assumptions, in practice one must consider all of the traditional concerns of one-dimensional spectral analysis (e.g., stationarity, end effects, aliasing, statistical consistency, window length, etc.).

3 Future Directions

Although the descriptive spatio-temporal methods described above will remain essential to the study of climate, other methods are being developed to accommodate the increasingly complex inference that is required to understand climate and its role in the earth system. For example, the problem of detecting anthropogenic climate change signals in the presence of natural climate variability and long-lead (seasonal and beyond) climate prediction are two areas that have led to recent

development of alternative statistical approaches. Particular problems include spatio-temporal nonstationarity, nonlinear prediction, non-Gaussian data, measurement error, change of spatial support, dimension reduction, signal-to-noise variability, and the incorporation of prior knowledge (qualitative or physical) into stochastic models. Statistical methods to address some of these issues include neural networks, wavelet analysis, multitaper spectral methods, singular spectrum analysis, total least squares analysis, hidden Markov models, non-Gaussian state-space models, and hierarchical Bayesian analysis. Only time will tell which of these (or others) will become essential tools of climatology. However, it is likely that as the inferential questions become increasingly refined, all of these methods (or variants) will play important roles.

Acknowledgments

The author is grateful to Professor D. Zimmerman for the invitation to participate in the EOLSS project. The writing of this manuscript was supported in part by NASA grant NAG5-8344.

Bibliography

- Bretherton C.S., Smith C., and Wallace J.M. (1992). An intercomparison of methods for finding coupled patterns in climate data. *Journal of Climate* **5**, 541-560. [This paper presents a comprehensive overview of various space-time methods used in climatology.]
- Daley, R. (1991). *Atmospheric Data Analysis*, Cambridge: Cambridge University Press. [This book presents essential aspects of meteorological data assimilation and optimal spatial prediction.]
- Freiberger W., and Grenander, U. (1965). On the formulation of statistical meteorology. *Review of the International Statistical Institute* **33**, 59-86. [A nice early review of statistics in meteorology.]
- Ghil M., Cohn S.E., Tavantzis J., Bube K., and Isaacson, E. (1981) Application of estimation theory to numerical weather prediction. *Dynamic Meteorology: Data Assimilation Methods*, L. Bengtsson, M. Ghil, and E. Källén, eds., New York: Springer-Verlag, 139-224. [Seminal paper describing the use of Kalman filters in the data assimilation process.]
- Haslett J. (1989) Space time modelling in meteorology - A review. *Bulletin*

- International Statistics Institute* **51**, 229-246. [A nice review of spatio-temporal statistical methods in meteorology, focused mainly on spatial prediction.]
- Hasselmann K. (1988) PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns. *Journal of Geophysical Research* **93**, 11015 - 11021. [Seminal work describing principal oscillation pattern analysis in a general context.]
- Hayashi Y. (1982) Space-time spectral analysis and its applications to atmospheric waves. *Journal Meteorological Society of Japan* **60**, 156-171. [Definitive paper on space-time spectral analysis and the assumptions relative to atmospheric waves.]
- Preisendorfer R.W. (1988) *Principal Component Analysis in Meteorology and Oceanography*, Amsterdam: Elsevier. [Definitive work on principal component analysis (EOFs) in the geosciences.]
- Thiébaux H.J., and Pedder M.A. (1987) *Spatial Objective Analysis with Applications in Atmospheric Science*, London: Academic Press. [Classic book describing optimal spatial prediction and data assimilation in the atmospheric sciences.]
- von Storch H., Bürger G., Schnur R., and von Storch J.-S. (1995) Principal oscillation patterns: a review. *Journal of Climate* **8**, 377-400. [Excellent review paper describing principal oscillation patterns.]
- von Storch H., and Zwiers F.W. (1999) *Statistical Analysis in Climate Research*, Cambridge: Cambridge University Press. [Definitive textbook on statistical methods in climatology.]
- Wikle C.K., Berliner L.M., and Cressie N. (1998) Hierarchical Bayesian space-time models, *Environmental and Ecological Statistics* **5**, 117-154. [Seminal paper describing hierarchical Bayesian spatio-temporal models in a climate application.]
- Wilks D.S. (1995) *Statistical Methods in the Atmospheric Sciences*, San Diego: Academic Press. [Excellent introductory textbook describing statistical methods used in atmospheric science.]