**Spatio-temporal statistical models
with applications to atmospheric processes**

by

Christopher Kim Wikle

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Departments: Statistics;

Geological and Atmospheric Science

Co-majors: Statistics;

Meteorology

Major Professors: Noel Cressie and Tsing-Chang Chen

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

Iowa State University

Ames, Iowa

1996

The Government reserves for itself and others acting on its behalf a royalty free, nonexclusive, irrevocable, world-wide license for governmental purposes to publish, distribute, translate, duplicate, exhibit, and perform any such data copyrighted by the contractor.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

**Portions of this document may be illegible
electronic image products. Images are
produced from the best available original
document.**

Graduate College
Iowa State University

This is to certify that the Doctoral dissertation of

Christopher K. Wikle

has met the dissertation requirements of Iowa State University

Co-major Professor

Co-major Professor

For the Co-major Department

For the Co-major Department

For the Graduate College

TABLE OF CONTENTS

GENERAL INTRODUCTION	1
1 Background	1
2 Dissertation Organization	4
2.1 Spatio-Temporal Statistical Methods in the Atmospheric Sciences	4
2.2 On the Semiannual Variation in the Northern Hemisphere Extratropical Height Field	4
2.3 Seasonal Variation of Lower Stratospheric Mixed Rossby-Gravity Waves over the Tropical Pacific	5
References	6
 SPATIO-TEMPORAL STATISTICAL METHODS	
IN THE ATMOSPHERIC SCIENCES	8
1 Introduction	8
1.1 Notation and General Model Assumptions	9
2 Empirical Orthogonal Function (EOF) Analysis	9
2.1 Continuous K-L Formulation	10
2.2 Discrete EOF Analysis	12
2.3 EOFs in the Presence of Measurement Error	14
2.4 EOFs as a Method of Smoothing	15
2.5 Estimation of EOFs	18
2.6 Variations on the Standard EOF Analysis	20
3 Principal Interaction Patterns (PIPs)	25

3.1	Application of PIPs	27
3.2	Alternate Views of PIPs	28
4	Principal Oscillation Patterns (POPs)	29
4.1	Formulation of POPs	29
4.2	Physical Implication of POPs	31
4.3	Estimation of POPs	33
4.4	Application of POPs	34
4.5	POPs on the EOF Basis	37
4.6	Extensions of POPs	37
5	Space-Time Canonical Correlation Analysis (CCA)	41
5.1	Two-Field Spatial-Temporal CCA	42
5.2	Estimation of CCA	45
5.3	Modifications of CCA	45
5.4	CCA in the Presence of Measurement Error	48
5.5	CCA in Continuous Space	50
6	Conclusion	51
	References	52
ON THE SEMIANNUAL VARIATION IN THE		
	NORTHERN HEMISPHERE EXTRATROPICAL HEIGHT FIELD	57
	Abstract	57
1	Introduction	58
2	Diagnostic Analysis	60
2.1	The Average Extratropical Northern Hemisphere Semiannual Oscillation	60
2.2	Northern Hemisphere Midlatitude Semiannual Oscillation	61
2.3	Northern Hemisphere High Latitude Semiannual Oscillation	63
2.4	Vertical Structure	64
3	Discussion	66

*Preprint - Removed for
separate
processing*

4	Conclusions	67
	Acknowledgements	68
	References	68
SEASONAL VARIATION OF LOWER STRATOSPHERIC		
MIXED ROSSBY-GRAVITY WAVES		
OVER THE TROPICAL PACIFIC 77		
	Abstract	77
1	Introduction	78
	1.1 MRGW Forcing	78
	1.2 Seasonal Variation of MRGWs	79
2	Data and Methods	82
	2.1 Data	82
	2.2 Methods	83
3	Analysis	84
	3.1 Identification of Mixed Rossby-Gravity Waves	84
	3.2 Seasonal Variation of MRGWs	86
	3.3 Cyclic Spectral Analysis	90
4	Discussion	91
5	Conclusions	95
	Acknowledgements	96
	Appendix A: Seasonally Varying Spectral Analysis (SVSA)	96
	Appendix B: Autoregressive Cyclic Spectral Analysis	97
	Appendix C: Stochastic 2-Channel AR Simulation	100
	References	101

A SPATIALLY DESCRIPTIVE, TEMPORALLY DYNAMIC

STATISTICAL MODEL WITH APPLICATIONS TO

ATMOSPHERIC PROCESSES 118

*Preprint - removed
for separate processing*

*Preprint - removed for
separate processing*

Abstract	118
1 Introduction	119
2 Statistical Model	123
2.1 Kalman Filter Representation	126
3 Estimation of Model Parameters	130
3.1 Estimation of Model Covariances	130
3.2 Estimation of State Matrix $\mathbf{B}$	134
3.3 Estimation of $\mathbf{C}_0^Z$, $\mathbf{C}_1^Z$, and $\mathbf{C}_0^Y$	134
4 Selection of Model Basis Functions	135
4.1 The EOF Basis Set	136
4.2 Implementation of the EOF Basis	138
5 Simulation Example	143
5.1 Description of Simulation	143
5.2 SDTD Model Results with Simulated Data	145
5.3 Comparison to Other Methods	148
6 South China Sea Precipitation Example	150
6.1 Data	150
6.2 Exploratory Data Analysis	151
6.3 Implementing the SDTD Model	153
6.4 SDTD Model Results with Precipitation Data	154
7 Conclusion	156
Acknowledgements	158
Appendix A: The Spatio-Temporal Kalman Filter	159
Appendix B: Simple Kriging in the Presence of Measurement Error	163
References	165
GENERAL CONCLUSION	183
ACKNOWLEDGEMENTS	187

GENERAL INTRODUCTION

1 Background

Perhaps, it could be argued, we humans have a fundamental need to understand the world in which we live. At least, it is safe to say that it is usually in our best interest to do so. One feature of our world where this is certainly true is the atmosphere, specifically its long-term (i.e., climate) and short-term (i.e., weather) variability. In fact, one doesn't have to look too far to see the impact of weather and climate in virtually every aspect of our daily lives. Atmospheric influences are evident in agriculture, commerce, travel, recreation, and so forth. Consequently, characterization of atmospheric variability is more than just a curiosity, it is essential.

Broadly, we might say that the atmosphere/ocean 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, we 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, our knowledge is incomplete at each of the observation, summarization, and inference stages, and thus our understanding of the atmosphere is clouded by uncertainty. Consequently, by the time we get to the prediction phase, our lack of certainty, combined with the nonlinear dynamics of the system, contributes to what is now known as *dynamical chaos*. As originally outlined by Lorenz(1963), chaos implies a fundamental lack of predictability. However, all is not lost, as over the last 100 years or so, the science of statistics has given us

many tools with which to evaluate, quantify, and exploit probabilistic uncertainty. Although we are always faced with the inherent chaotic nature of the atmosphere/ocean system, we can approach many of the relevant scientific questions from a probabilistic viewpoint, which allows us to make useful inferences in the presence of uncertainty, at least for relatively large spatial scales and relatively short temporal scales. Furthermore, we are then able to look for possible associations within and between variables in the system, which may allow us to extend our 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 (e.g., Chatfield 1989, p. 126) and the "nugget effect" in geostatistics (e.g., Cressie 1993, p. 59). 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 (GDP) is relatively large. Thus, the location of a measuring site and its temporal sampling frequency may have very little to do with science. Therefore, to gain scientific insight, we must consider these uncertainties when framing our questions, choosing our analysis techniques, and interpreting our 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 statistical methods courses (which assume independent and identically distributed data) do not apply.

Traditionally, researchers in the atmospheric and related sciences have generally focused on

relatively simple time-series approaches and, outside of the data assimilation area of speciality, simple descriptive spatial techniques. These approaches have proven very valuable. However, as our already enormous data sets keep growing due to new observing platforms (e.g., satellite, radar, lidar, and profiler data), and as we ask more penetrating scientific questions with possibly severe implications (e.g., "given the observations to date, is there evidence of anthropogenic climate change?"), we must have more sophisticated techniques with which to handle the uncertainty in the data. Atmospheric scientists have risen to the challenge in recent years, and have exploited spatio-temporal methods such as Empirical Orthogonal Functions or EOFs (e.g., Lorenz 1956; Preisendorfer 1988), spatio-temporal Canonical Correlation Analysis or CCA (e.g., Glahn 1968; Bretherton et al. 1992), and Principal Oscillation Patterns or POPs (e.g., Hasselmann 1988; von Storch et al. 1988,1994). These are excellent tools with which to summarize data and have, to a lesser extent, sometimes been considered for prediction. However, often the assumptions under which these methods should be used are either poorly understood or ignored. Furthermore, as typically used, such methods contain no mechanism for spatial prediction. This is unfortunate since incomplete spatial sampling of the atmosphere makes spatial prediction an issue of critical importance. It is apparent that there is a need for additional spatio-temporal methods in the atmospheric sciences.

There have been some notable recent attempts to systematically introduce statistical ideas to the atmospheric science community (e.g., Thiébaux 1994; Wilks 1995), although the scope of these works is relatively broad. My goal in this dissertation is to continue to narrow the gap between the need for atmospheric science related spatio-temporal methods, and modern statistical approaches to such methods. To do this, I will first provide an overview of spatio-temporal statistical methods commonly used in the atmospheric sciences, from a statistical point of view. Next, I will use statistical techniques ranging from the very simple (e.g., harmonic analysis) to "state-of-the-art" (e.g., autoregressive cyclic spectral analysis) to characterize the variability of outstanding atmospheric science problems concerned with the spatial structure of the extratropical semi-annual cycle, and temporal variability of mixed-Rossby gravity waves over the

tropical Pacific. Finally, I will develop and implement a new and very general spatio-temporal statistical model that is both spatially descriptive and temporally dynamic. This model is applied to monthly precipitation data over the South China Sea.

2 Dissertation Organization

The dissertation is organized according to a "paper format". Following the General Introduction, a preliminary chapter is included to provide an overview of spatio-temporal statistical methods in the atmospheric sciences. This is then followed by three studies which have been, or will be, submitted for publication in an appropriate atmospheric science or statistics journal. Each paper is self-contained and includes a full literature review. Since each paper is independent, equation numbers only apply to the paper at hand. The papers are then followed by a general conclusion. A brief summary of each chapter is described below.

2.1 Spatio-Temporal Statistical Methods in the Atmospheric Sciences

In this overview, I examine the traditional spatio-temporal statistical methods used in the atmospheric sciences. These methods are considered from a statistical perspective. Although this section is primarily a review, many of the statistical issues that are considered have not, to my knowledge, been considered in the context of these methods. As a consequence, several "open questions" are posed in this review.

2.2 On the Semiannual Variation in the Northern Hemisphere Extratropical Height Field

Based upon the application to atmospheric data of a recent signal-processing technique for identifying periodic components in the presence of unknown noise (Wikle et al. 1995), it became clear that there is a distinct semiannual oscillation (SAO) in most atmospheric variables. It was also clear that the strength of this SAO signal varies depending on spatial location. Since no unified view of the extratropical Northern Hemisphere (NH) SAO has been published, I

sought to find a mechanism by which the SAO spatial variation could be characterized. Along with Professor T.-C. Chen, I discovered that the NH midlatitude SAO in 500hPa geopotential height could be explained almost entirely as a result of spatial and temporal asymmetries in the annual variation of the stationary eddies. We then concluded that the mechanism for the SAO in the NH is simply a result of land-sea contrasts, similar to the mechanism proposed by van Loon (1967) for the Southern Hemisphere (SH) SAO. There is a fundamental difference, however, in that the NH extratropics are dominated by *east-west* land-sea contrasts due to the large continental land masses in the NH, while the SH land-sea contrast reflects the *north-south* differential heating between Antarctica and the surrounding ocean. These results are important primarily for their application to sensitivity tests of atmospheric general circulation models (GCMs). In particular, before we can accept GCM results related to interannual and interdecadal variability, we must be sure that they can properly simulate the relatively simple components of the atmospheric seasonal cycle, namely the SAO. To date, they are *not* able to do this.

2.3 Seasonal Variation of Lower Stratospheric Mixed Rossby-Gravity Waves over the Tropical Pacific

It has long been understood that the atmospheric general circulation is principally driven by diabatic heating with its centers located in the tropics. This diabatic heating is mostly attributable to the latent heat released from cumulus convection. Moreover, it has been shown that tropical planetary waves can be modulated by this diabatic heating. Thus, since tropical convective activity is known to show semiannual variability (e.g., Chen and Wu 1992), we would expect to see a semiannual signal in equatorial wave activity.

Along with Dr. R. Madden of the National Center for Atmospheric Research (NCAR) and Professor T.-C. Chen, I examined the seasonal variability of mixed Rossby-gravity waves (MRGWs) in lower stratospheric wind measurements over the equatorial Pacific. Utilizing the seasonally varying spectral analysis procedure developed by Madden (1986) and the recently developed autoregressive cyclic spectral analysis technique (e.g., Sherman and White 1995), I

found that there are significant twice-yearly peaks in MRGW activity, with peaks occurring in winter-early spring and in summer-early fall. In addition, the seasonally varying spectral analysis suggested that there is a previously unknown convergence of horizontal momentum flux associated with these waves, and that the sign of that convergence is different during the times of the two seasonal maxima. There was also an indication from the cyclic spectral analysis that the frequency of the MRGWs may be different during the two maxima.

2.3.1 A Spatially Descriptive, Temporally Dynamic Statistical Model with Applications to Atmospheric Processes

In this paper, with Professor N. Cressie, a new spatio-temporal statistical model is proposed that attempts to consider the influence of both temporal and spatial variability. Unlike the traditional spatio-temporal methods used in the atmospheric science literature (and outlined in the first chapter of this dissertation), this method is mainly concerned with *prediction* in space and time. Our predictive model is temporally dynamic in that it exploits the unidirectional flow of time in an autoregressive framework. In addition, the model is spatially descriptive in the sense that although spatial correlation is modeled by a spatially colored noise process, no causative interpretation is associated with this noise. With the inclusion of measurement error, this formulation naturally leads to the development of a spatio-temporal Kalman filter. We can then use this Kalman filter to predict at future times and at locations for which we do not have data. We demonstrate this method through simulation and show its utility to atmospheric science by using it to predict monthly precipitation throughout the data-sparse South China Sea region.

References

- Bretherton, C.S., C. Smith, and J.M. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate data. *J. Climate*, **5**, 541-560.

- Chatfield, C., 1989: *The Analysis of Time Series: An Introduction*, Fourth Edition, Chapman and Hall, 241pp.
- Chen, T.-C., and K.D. Wu, 1992: Semi-annual oscillation of the global divergent circulation. *Tellus*, **44A**, 357-365.
- Cressie, N.A.C., 1993: *Statistics for Spatial Data, Revised Edition*, Wiley, 900pp.
- Glahn, H.R., 1968: Canonical correlation and its relationship to discriminant analysis and multiple regression. *J. Atmos. Sci.*, **25**, 23-31.
- Hasselmann, K., 1988: PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.*, **93**, 11015 - 11021.
- Lorenz, E.N., 1956: Empirical orthogonal functions and statistical weather prediction. *Sci. Rept. No. 1, Statistical Forecasting Project*, MIT, 49 pp.
- Lorenz, E.N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130-141.
- Madden, R.A., 1986: Seasonal variations of the 40-50 day oscillation in the tropics. *J. Atmos. Sci.*, **43**, 3138-3158.
- Preisendorfer, R.W., 1988: *Principal Component Analysis in Meteorology and Oceanography*, Elsevier, 425 pp.
- Sherman, P.J., and L.B. White, 1995: Improved periodic spectral analysis with application to diesel vibration data. *J. Acoust. Soc. Amer.*, **98**, 3285-3301.
- Thiébaux, H.J., 1994: *Statistical Data Analysis for Ocean and Atmospheric Sciences*, Academic Press, 247pp.
- van Loon, H., 1967: The half-yearly oscillation in middle and high southern latitudes and the coreless winter. *J. Atmos. Sci.*, **24**, 472-486.
- von Storch, H., G. Bürger, R. Schnur, and J.-S. von Storch, 1995: Principal oscillation patterns: a review. *J. Climate*, **8**, 377-400.
- von Storch, H., T. Bruns, I. Fischer-Bruns, and K. Hasselmann, 1988: Principal oscillation pattern analysis of the 30- to 60-day oscillation in a general circulation model equatorial troposphere. *J. Geophys. Res.*, **93**, 11022-11036.
- Wilks, D.S., 1995: *Statistical Methods in the Atmospheric Sciences*, Academic Press, 467pp.
- Wikle, C.K., P.J. Sherman, and T.-C. Chen, 1995: Identifying periodic components in atmospheric data using a family of minimum variance spectral estimators. *J. Climate*, **8**, 2352-2363.

SPATIO-TEMPORAL STATISTICAL METHODS IN THE ATMOSPHERIC SCIENCES

1 Introduction

Since virtually all meteorological and climatological processes involve variability over both space and time, it is imperative that statistical models for these processes also consider spatio-temporal variability. The atmospheric science literature contains many examples of statistical methods that capture various forms of spatio-temporal variability, both in diagnostic and prognostic applications. These methods include Empirical Orthogonal Function (EOF), Principal Interaction Pattern (PIP), Principal Oscillation Pattern (POP), and Canonical Correlation Analysis (CCA) techniques. There are often underlying assumptions and approximations associated with the application of these techniques that do not seem to be recognized or, at least, they are not reported. In particular, issues such as discrete vs. continuous space and time, prognostic vs. diagnostic application, measurement error vs. no measurement error, and the statistical criterion against which the methodology under consideration is optimal are not always considered explicitly.

My goal in this review is to examine the traditional spatio-temporal statistical methods used in the atmospheric sciences from a more statistical perspective. In addition, I shall try to emphasize the practical scientific motivation for these methods as they are (or should be) applied in the atmospheric sciences. Many of the statistical issues that are considered have not, to my knowledge, been considered in the literature. Thus, there may be several "open questions" that are posed along the way.

1.1 Notation and General Model Assumptions

Assume that we are given observations of a spatio-temporal process $Z(\mathbf{s}; t)$ at spatial locations $\mathbf{s} \in \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n\}$ in some spatial domain D_z and time $t \in \{1, 2, \dots, T\}$, where D_z is assumed to be two-dimensional Euclidean space unless otherwise noted. We may also have observations of some process $X(\mathbf{r}; t)$ at spatial locations $\mathbf{r} \in \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_m\} \subset D_x$ and time $t \in \{1, 2, \dots, T\}$, where D_x is some spatial domain such that D_z and D_x may overlap, depending on the application. In some cases we will assume that there is measurement error associated with the observation of these processes.

2 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 or PCA (e.g., Hotelling 1933), while in the continuous framework, it is simply a Karhunen-Loève (K-L) expansion (e.g., Loève 1963). Depending on the application, EOFs are usually used:

- 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
- to reduce the degrees of freedom (spatially) in large geophysical data sets while simultaneously reducing noise.

An extensive review of EOFs, their theory, and their application to meteorology and oceanography is contained in Preisendorfer (1988).

One finds in the meteorological literature, extensive use of EOFs since their introduction by Lorenz (1956). For example, they have been used for describing climate (e.g., Kutzbach 1967; Barnett 1977), for comparing simulations of general circulation models (e.g., Preisendorfer and Barnett 1983), for developing regression forecast techniques (e.g., Peagle and Haslam 1982), in weather classification (e.g., Christensen and Bryson 1966), in map typing (e.g., Richman 1981),

in the interpretation of geophysical fields (Olbed and Creutin 1986), and in the simulation of random fields, particularly non-homogeneous processes (Braud and Obled 1991). In addition, as in the psychometric literature for PCAs and as advocated by Richman (1981,1986) and others, 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 by Wallace and Dickinson (1972) and was shown to be very useful in applications to certain meteorological problems (e.g., Wallace 1972; Barnett 1983; Horel 1984).

In the remainder of this section, I will briefly present the K-L expansion, show how the discrete EOF formulation can be derived in a PCA context, and present a couple of issues related to EOFs and measurement error that have not been considered in the literature. Finally, I will briefly consider the central idea behind some variations of EOF analysis, namely, complex EOF analysis, multivariate EOF analysis, and extended EOF analysis.

2.1 Continuous K-L Formulation

We first consider a continuous spatial process measured at discrete time intervals. Our goal is to find an optimal and separable orthogonal decomposition of a spatio-temporal process $Z(\mathbf{s}; t)$. That is, we want

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

such that the $\text{var}(a_1(t)) > \text{var}(a_2(t)) > \dots$, 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 (e.g., see Papoulis 1965, p. 457-461). Suppose,

$$E[Z(\mathbf{s}; t)] = 0, \quad (2)$$

and define the covariance function as

$$E[Z(\mathbf{s}; t)Z(\mathbf{r}; t)] \equiv c_0^Z(\mathbf{s}, \mathbf{r}), \quad (3)$$

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:

$$c_0^Z(\mathbf{s}, \mathbf{r}) = \sum_{k=1}^{\infty} \lambda_k \phi_k(\mathbf{s}) \phi_k(\mathbf{r}), \quad (4)$$

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

$$\int_D c_0^Z(\mathbf{s}, \mathbf{r}) \phi_k(\mathbf{s}) d\mathbf{s} = \lambda_k \phi_k(\mathbf{r}), \quad (5)$$

where

$$\int_D \phi_k(\mathbf{s}) \phi_l(\mathbf{s}) d\mathbf{s} = \begin{cases} 1 & \text{for } k = l, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

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

$$Z(\mathbf{s}; t) = \sum_{k=1}^{\infty} a_k(t) \phi_k(\mathbf{s}), \quad (7)$$

where we call $\{\phi_k(\mathbf{s}) : \mathbf{s} \in D\}$ the k -th EOF and often refer to the associated time series $a_k(t)$ as 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,

$$a_k(t) = \int_D Z(\mathbf{s}; t) \phi_k(\mathbf{s}) d\mathbf{s}. \quad (8)$$

It is easy to verify that these time series are uncorrelated, with variance equal to the corresponding eigenvalues; that is,

$$E[a_i(t) a_k(t)] = \delta_{ik} \lambda_k, \quad (9)$$

where δ_{ik} equals one when $i = k$, and equals zero otherwise.

If we truncate the expansion (7) at K , yielding

$$Z_K(\mathbf{s}; t) \equiv \sum_{k=1}^K a_k(t) \phi_k(\mathbf{s}), \quad (10)$$

then it can be shown (e.g., Freiberger and Grenander 1965; Davis 1976) 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 we must solve numerically the Fredholm integral equation (5) to obtain the EOF basis functions. Cohen and Jones (1986) and Buell (1972,1975) give numerical quadrature solutions to this problem. The 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 $\{s_1, \dots, s_n\}$ for which there are data. Obled and Creutin (1986) have presented an elegant method for discretizing the K-L integral equation and for interpolating the eigenfunctions to locations where data are not available. See the third paper of this dissertation for an example of this approach.

2.2 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. Let us consider a discrete EOF analysis by using the PCA formulation as given in standard multivariate statistics books (e.g., Johnson and Wichern 1992), but according to the spatio-temporal notation we have introduced. In that case, let

$$\mathbf{Z}(t) \equiv (Z(s_1; t), \dots, Z(s_n; t))' \quad (11)$$

and define the k -th EOF ($k = 1, \dots, n$) to be

$$\boldsymbol{\psi}_k \equiv (\psi_k(s_1), \dots, \psi_k(s_n))', \quad (12)$$

where $\boldsymbol{\psi}_k$ is the vector in the linear combination

$$a_k(t) = \boldsymbol{\psi}_k' \mathbf{Z}(t). \quad (13)$$

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

$$\psi'_k \psi_k = 1 \quad (14)$$

and

$$\text{cov}(a_k(t), a_j(t)) = 0, \text{ for all } k \neq j. \quad (15)$$

This is equivalent to solving the eigensystem

$$\mathbf{C}_0^Z \Psi = \Psi \Lambda, \quad (16)$$

where

$$\mathbf{C}_0^Z \equiv E[\mathbf{Z}(t)\mathbf{Z}(t)'] \quad (17)$$

$$\Psi = (\psi_1, \dots, \psi_n) \quad (18)$$

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n), \quad (19)$$

and where

$$\text{var}(a_i(t)) = \lambda_i, \quad i = 1, \dots, n. \quad (20)$$

Then the solution of (1.16) is obtained by a symmetric decomposition of $\mathbf{C}_0^Z$,

$$\mathbf{C}_0^Z = \Psi \Lambda \Psi', \quad (21)$$

which is the PCA formulation.

It is straightforward to show (e.g., Cohen and Jones 1969; Buell 1972) 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. For a discussion of the effects of ignoring this distinction, see Karl et al. (1982).

2.3 EOFs in the Presence of Measurement Error

The EOF decomposition in the presence of measurement error has not been considered in the literature. We assume that the process $Z(\cdot; \cdot)$ is the sum of a smooth signal of interest Y and additive noise,

$$\mathbf{Z}(t) = \mathbf{Y}(t) + \boldsymbol{\epsilon}(t), \quad (22)$$

where

$$\mathbf{Y}(t) \equiv (Y(\mathbf{s}_1; t), \dots, Y(\mathbf{s}_n; t))' \quad (23)$$

$$\boldsymbol{\epsilon}(t) \equiv (\epsilon(\mathbf{s}_1; t), \dots, \epsilon(\mathbf{s}_n; t))'. \quad (24)$$

Taking the variance of $\mathbf{Z}(t)$ gives

$$\mathbf{C}_0^Z = \mathbf{C}_0^Y + \mathbf{C}_0^\epsilon, \quad (25)$$

where

$$\mathbf{C}_0^Y \equiv \text{var}[\mathbf{Y}(t)] \quad (26)$$

$$\mathbf{C}_0^\epsilon \equiv \text{var}[\boldsymbol{\epsilon}(t)]. \quad (27)$$

Now, assuming regularly spaced observations, we can show that

$$\boldsymbol{\Psi} \boldsymbol{\Lambda} \boldsymbol{\Psi}' = \mathbf{C}_0^Y + \mathbf{C}_0^\epsilon. \quad (28)$$

If we let the measurement error be white noise with variance σ_ϵ^2 , then

$$\mathbf{C}_0^\epsilon = \sigma_\epsilon^2 \mathbf{I} = \sigma_\epsilon^2 \boldsymbol{\Psi} \boldsymbol{\Psi}', \quad (29)$$

and

$$\mathbf{C}_0^Y = \boldsymbol{\Psi} (\boldsymbol{\Lambda} - \sigma_\epsilon^2 \mathbf{I}) \boldsymbol{\Psi}'. \quad (30)$$

Now, if we let

$$Y(\mathbf{s}; t) = \sum_{k=1}^n a_k^*(t) \psi_k(\mathbf{s}), \quad (31)$$

where

$$a_k^*(t) \equiv \boldsymbol{\psi}_k' \mathbf{Y}(t). \quad (32)$$

Then,

$$E[\mathbf{a}^*(t)\mathbf{a}^*(t)'] = \mathbf{\Lambda} - \sigma_\epsilon^2 \mathbf{I}, \quad (33)$$

where

$$\mathbf{a}^*(t) \equiv (a_1^*(t), \dots, a_n^*(t))'. \quad (34)$$

Thus, the noisy Z process and the smooth Y process have the same eigenfunctions $\{\psi_k\}$, but different eigenvalues (i.e., $\{\lambda_k\}$ and $\{\lambda_k - \sigma_\epsilon^2\}$, respectively). In addition, the amplitude time series for both processes have the same mean (zero), but the variance of the Y process amplitude time series amplitude is less (as expected). Thus, if white measurement error is present (and it always is with observational data), then the spatial EOF spatial patterns (i.e., the $\{\psi_k\}$) are not affected, but the variance explained by each pattern must be adjusted accordingly. This is seldom (if ever) done in practice.

2.4 EOFs as a Method of Smoothing

The truncated EOF expansion in the discrete framework is given by

$$Z_K(\mathbf{s}; t) = \sum_{k=1}^K a_k(t) \psi_k(\mathbf{s}). \quad (35)$$

As was the case for the EOF truncation in continuous space (10), the EOF basis can be shown to minimize the variance of the truncation error (e.g., Davis 1976). In practice, it is common to assume that this truncation can be used as a prediction of the smooth process Y . That is,

$$\hat{Y}(\mathbf{s}; t) \equiv Z_K(\mathbf{s}; t). \quad (36)$$

Although not considered in the literature, it is natural to ask if this is an optimal predictor.

To investigate the optimality of (36), we consider the space-only process analogous to (22), that is,

$$Z(\mathbf{s}) = Y(\mathbf{s}) + \epsilon(\mathbf{s}). \quad (37)$$

Assume we have data at n locations $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ and that we are interested in predicting $Y(\mathbf{s}_i)$ from $\mathbf{Z} \equiv (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))'$. If we assume that $Z(\cdot)$ is Gaussian, then the optimal predictor

is just a linear predictor. After Cressie (1993, p. 110), this optimal predictor is given by

$$E[Y(\mathbf{s}_i) | \mathbf{Z}] = \mathbf{c}^Y(\mathbf{s}_i)'[\mathbf{C}^Z]^{-1}\mathbf{Z}; \quad i = 1, \dots, n, \quad (38)$$

where we have assumed that the means of $\mathbf{Z}$ and $Y(\mathbf{s}_i)$ are known to be zero, and we define

$$\mathbf{c}^Y(\mathbf{s}_i) \equiv \text{cov}(Y(\mathbf{s}_i), \mathbf{Y}) \quad (39)$$

$$\mathbf{C}^Z \equiv \text{var}(\mathbf{Z}) \quad (40)$$

and

$$\mathbf{Y} \equiv (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))'. \quad (41)$$

Now, if we let $E[\epsilon(\mathbf{s}_i)\epsilon(\mathbf{s}_k)] = \sigma_\epsilon^2$ for $i = k$, and equal to zero otherwise, then we can write

$$\mathbf{C}^Z = \mathbf{C}^Y + \sigma_\epsilon^2 \mathbf{I}, \quad (42)$$

where

$$\mathbf{C} \equiv \text{var}(\mathbf{Y}). \quad (43)$$

Now, since $\mathbf{C}^Y$ is real and symmetric, we can write the following decomposition:

$$\mathbf{C}^Y = \Psi \Lambda^* \Psi', \quad (44)$$

where the eigenvectors and eigenvalues are:

$$\Psi \equiv (\psi_1, \dots, \psi_n) \quad (45)$$

$$\psi_k \equiv (\psi_k(\mathbf{s}_1), \dots, \psi_k(\mathbf{s}_n))', \quad k = 1, \dots, n \quad (46)$$

$$\Lambda^* \equiv \text{diag}(\lambda_1^*, \dots, \lambda_n^*). \quad (47)$$

Then, using the orthogonality $\Psi\Psi' = \Psi'\Psi = \mathbf{I}$, we obtain

$$\mathbf{C}^Z = \Psi \Lambda^* \Psi' + \sigma_\epsilon^2 \Psi \Psi' \quad (48)$$

$$= \Psi \Lambda \Psi', \quad (49)$$

where

$$\Lambda = \Lambda^* + \sigma_\epsilon^2 \mathbf{I}. \quad (50)$$

Now, (38) can be written

$$E[Y(\mathbf{s}_i) | \mathbf{Z}] = \mathbf{c}^Y(\mathbf{s}_i)' \Psi \Lambda^{-1} \Psi' \mathbf{Z}. \quad (51)$$

Furthermore, since

$$E[Y(\mathbf{s}_i)Y(\mathbf{s}_j)] = \sum_{k=1}^n \lambda_k^* \psi_k(\mathbf{s}_i) \psi_k(\mathbf{s}_j), \quad (52)$$

then

$$\mathbf{c}^Y(\mathbf{s}_i) = \Psi \Lambda^* \boldsymbol{\psi}(\mathbf{s}_i), \quad (53)$$

where

$$\boldsymbol{\psi}(\mathbf{s}_i) \equiv (\psi_1(\mathbf{s}_i), \dots, \psi_n(\mathbf{s}_i))'; \quad i = 1, \dots, n. \quad (54)$$

In addition, as for the EOF case,

$$\mathbf{Z} = \Psi \mathbf{a}, \quad (55)$$

where

$$\mathbf{a} \equiv (a_1, \dots, a_n)', \quad (56)$$

and

$$a_k = \boldsymbol{\psi}'_k \mathbf{Z}. \quad (57)$$

We can then write the optimal predictor as

$$E[Y(\mathbf{s}_i) | \mathbf{Z}] = \boldsymbol{\psi}(\mathbf{s}_i) \Lambda^* \Psi' \Psi \Lambda^{-1} \Psi' \Psi \mathbf{a} \quad (58)$$

$$= \sum_{k=1}^n \frac{\lambda_k^*}{\lambda_k^* + \sigma_\epsilon^2} a_k \psi_k(\mathbf{s}_i). \quad (59)$$

Now, the truncated predictor of interest is given by

$$Z_K(\mathbf{s}_i) = \sum_{k=1}^K a_k \psi_k(\mathbf{s}_i); \quad i = 1, \dots, n. \quad (60)$$

Thus, the optimal predictor (59) is equivalent to the truncated predictor (60) in the trivial case where $\sigma_\epsilon^2 = 0$ and $\sum_{l=K+1}^n a_l \psi_l(\mathbf{s}_i) = 0$. In general, setting (59) equal to (60) gives

$$\sum_{k=1}^K \frac{\lambda_k^*}{\lambda_k^* + \sigma_\epsilon^2} a_k \psi_k(\mathbf{s}_i) + \sum_{l=K+1}^n \frac{\lambda_l^*}{\lambda_l^* + \sigma_\epsilon^2} a_l \psi_l(\mathbf{s}_i) = \sum_{k=1}^K a_k \psi_k(\mathbf{s}_i), \quad (61)$$

which results in the equality constraint

$$\sum_{k=1}^K \left(\frac{\lambda_k^*}{\lambda_k^* + \sigma_\epsilon^2} - 1 \right) a_k \psi_k(\mathbf{s}_i) = - \sum_{l=K+1}^n \frac{\lambda_l^*}{\lambda_l^* + \sigma_\epsilon^2} a_l \psi_l(\mathbf{s}_i) ; i = 1, \dots, n. \quad (62)$$

Clearly, (62) shows that if $\sigma_\epsilon^2 = 0$, then optimality is achieved if $a_l = 0$ or $\lambda_l^* = 0$ for $l = K + 1, \dots, n$. Thus, in the absence of measurement error, if the truncation parameter K is large so that the λ_l^* are very close to zero, then the truncated EOF smoothing technique is "nearly" optimal. However, in the presence of measurement error, the discrete orthogonality of the eigenvectors in (62) implies optimality when $a_k = 0$ for $k = 1, \dots, K$ and $\lambda_l^* = 0$ or $a_l = 0$ for $l = K + 1, \dots, n$. Under such conditions, one would probably not be interested in an EOF analysis. Thus, we can state that $Z_K(\mathbf{s}_i)$ is *not in general* an optimal predictor of $Y(\mathbf{s}_i)$. Consequently, the smoothed quantity $Z_K(\mathbf{s}_i)$ should be viewed as an *ad hoc* method for removing noise.

2.5 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, we need an estimate of

$$\mathbf{C}_0^Z = E[\mathbf{Z}(t)\mathbf{Z}(t)'], \quad (63)$$

where $\mathbf{Z}(t)$ is assumed to have zero mean. The MOM estimator for an element of $\mathbf{C}_0^Z$,

$$c_0^Z(\mathbf{s}_i, \mathbf{s}_j) \equiv E(Z(\mathbf{s}_i; t)Z(\mathbf{s}_j; t)), \quad \forall t, \quad (64)$$

is given by

$$\hat{c}_0^Z(\mathbf{s}_i, \mathbf{s}_j) \equiv (1/T) \sum_{t=1}^T [Z(\mathbf{s}_i; t) - \hat{\mu}_Z(\mathbf{s}_i; t)][Z(\mathbf{s}_j; t) - \hat{\mu}_Z(\mathbf{s}_j; t)], \quad (65)$$

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 quite typically show a nonzero mean. Possible choices of $\hat{\mu}_Z(\mathbf{s}_i; t)$

include the time mean,

$$\hat{\mu}_Z(\mathbf{s}_i) \equiv (1/T) \sum_{t=1}^T Z(\mathbf{s}_i; t), \quad (66)$$

the space mean,

$$\hat{\mu}_Z(t) \equiv (1/n) \sum_{i=1}^n Z(\mathbf{s}_i; t), \quad (67)$$

and the grand mean,

$$\hat{\mu}_Z \equiv (1/nT) \sum_{t=1}^T \sum_{i=1}^n Z(\mathbf{s}_i; t). \quad (68)$$

To the best of my knowledge, the investigation of the proper choice for estimating the mean has not appeared in the literature. Typically, investigators use the time mean (66), but it is not at all clear that this is the best choice. Further investigation is needed. Perhaps, until a study of the effect of the different choices can be performed, the best thing is to use the grand mean (68).

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

$$\hat{\mathbf{C}}_0^Z = \hat{\Psi} \hat{\Lambda} \hat{\Psi}'. \quad (69)$$

Note that Lawley (1956) derived approximate formulas for the bias and variance of the standard eigenvalue estimator. Later, von Storch and Hannoschock (1985) extended Lawley's results and considered the bias and variance in the EOF case. Based on theory and Monte Carlo simulation, they found that the sample eigenvalue $\hat{\lambda}_k$ is a biased estimator of λ_k . This bias is positive for the larger λ_k 's and negative for the smaller λ_k 's. They note that unbiased estimators can be constructed, but that the decrease in bias is accompanied by an increase in the variance of the estimator.

Furthermore, it has been shown that the sampling error associated with the estimated EOFs leads to numerical instability in the eigenvectors (e.g., Gray 1981; North et al. 1982). This has led to sampling-based selection strategies for the truncation level, K . Many of these are described in Preisendorfer (1988).

2.6 Variations on the Standard EOF Analysis

In this section, I shall briefly examine the idea behind several extensions of the standard EOF analysis described above.

2.6.1 Complex EOF Analysis

Consider a spatio-temporal process consisting of a sinusoid in one spatial dimension that is invariant in time:

$$Z(s; t) = B \sin(ls), \quad (70)$$

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):

$$Z(s; t) = B \sin(ls + \omega t) \quad (71)$$

$$= B \cos(\omega t) \sin(ls) - B \sin(\omega t) \cos(ls), \quad (72)$$

where ω is the temporal frequency. Thus, as the difference between (70) and (72) clearly shows, in order to characterize the phase propagation of such a sinusoid, we need information 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 (e.g., Chatfield 1989).

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 (analogous to equation (70) above). To extend the EOF analysis to the study of spatial structures that can propagate in time (e.g., analogous to equation (71) above), Wallace and Dickinson (1972) developed 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. It was originally presented by Rasmusson et al. (1981), and has since been used by a number of investigators (e.g., Barnett 1983; Trenberth and Shin 1984). 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 (72), the original data and its Hilbert transform allow the examination of propagating disturbances. Horel (1984) gives an excellent discussion of some of the theoretical and practical issues related to CEOF analysis. In our development, the method is described following the approach of Barnett (1983).

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

$$Z(\mathbf{s}_j; t) = \sum_{\omega} \alpha_j(\omega) \cos(\omega t) + \beta_j(\omega) \sin(\omega t), \quad (73)$$

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:

$$Z^f(\mathbf{s}_j; t) = \sum_{\omega} \gamma_j(\omega) e^{-i\omega t}, \quad (74)$$

where $\gamma_j(\omega) = \alpha_j(\omega) + i\beta_j(\omega)$. Using the definition of $\gamma_j(\omega)$, we can expand (74) to obtain:

$$Z^f(\mathbf{s}_j; t) = Z(\mathbf{s}_j; t) + i\tilde{Z}(\mathbf{s}_j; t), \quad (75)$$

where

$$Z(\mathbf{s}_j; t) = \alpha_j(\omega) \cos(\omega t) + \beta_j(\omega) \sin(\omega t), \quad (76)$$

and,

$$\tilde{Z}(\mathbf{s}_j; t) = \beta_j(\omega) \cos(\omega t) - \alpha_j(\omega) \sin(\omega t). \quad (77)$$

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 $\frac{\pi}{2}$. Barnett (1983) gives several methods for estimating the Hilbert transform.

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

$$\mathbf{C}_0^{Zf} = [c_0^{Zf}(\mathbf{s}_j, \mathbf{s}_k)]_{j,k=1,\dots,n} \quad (78)$$

where

$$c_0^{Zf}(\mathbf{s}_j, \mathbf{s}_k) \equiv E[Z^f(\mathbf{s}_j; t)^* Z^f(\mathbf{s}_k; t)], \quad (79)$$

and where $*$ denotes the complex conjugate. Note that $\mathbf{C}_0^{Zf}$ 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 we are only interested in phenomena occurring over a certain spectral frequency range of ω , then we can filter accordingly the original process $Z(\cdot; t)$ and its Hilbert transform $\tilde{Z}(\cdot; t)$ before the CEOF analysis.

Since $\mathbf{C}_0^{Zf}$ is Hermitian, it possesses real eigenvalues $\{\lambda_k\}$ and complex eigenvectors,

$$\gamma_k \equiv (\gamma_k(\mathbf{s}_1), \dots, \gamma_k(\mathbf{s}_n))'; \quad k = 1, \dots, n. \quad (80)$$

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:

$$Z^f(\mathbf{s}_i; t) = \sum_{k=1}^n a_k(t) \gamma_k^*(\mathbf{s}_i), \quad (81)$$

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

$$a_k(t) = \sum_{i=1}^n Z^f(\mathbf{s}_i; t) \gamma_k(\mathbf{s}_i). \quad (82)$$

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

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

$$\theta_k(\mathbf{s}_i) = \arctan \left[\frac{\text{Im}(\gamma_k(\mathbf{s}_i))}{\text{Re}(\gamma_k(\mathbf{s}_i))} \right]. \quad (83)$$

This function can take any value between $-\pi$ and π . In the case of the simple sinusoid with temporal phase (71), this corresponds to ls . In that case the spatial phase will go through one complete cycle (2π) over the distance $2\pi/l$. It should be noted that for data fields which include many different scales of variability, the spatial phase plot can be very difficult to interpret. The pre-filtering procedure described above generally improves interpretability.

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

$$S_k(\mathbf{s}_i) = [\gamma_k(\mathbf{s}_i)\gamma_k^*(\mathbf{s}_i)]^{1/2}. \quad (84)$$

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:

$$\xi_k(t) = \arctan \left[\frac{\text{Im}(a_k(t))}{\text{Re}(a_k(t))} \right]. \quad (85)$$

Consider the simple sinusoid example in (71). 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:

$$R_k(t) = [a_k(t)a_k^*(t)]^{1/2}. \quad (86)$$

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

2.6.2 Multivariate EOF Analysis

Often, we may be interested in the simultaneous analysis of two or more processes. Kutzbach (1967) used a form of EOF analysis which simultaneously considered several meteorological variables at many spatial locations and times. Preisendorfer (1988) considers this approach at length. A brief description of the basic idea behind this multivariate EOF analysis methodology follows.

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, we can write

$$\mathbf{W}(t) \equiv [\mathbf{Z}(t)' \mathbf{X}(t)']', \quad (87)$$

where

$$\mathbf{X}(t) \equiv (X(\mathbf{s}_1; t), \dots, X(\mathbf{s}_n; t))', \quad (88)$$

and $\mathbf{Z}(t)$ is defined in (11). Then, the covariance matrix of $\mathbf{W}(t)$ is given by

$$\mathbf{C}_0^W = E[\mathbf{W}(t) \mathbf{W}(t)']. \quad (89)$$

Thus, it is clear that this matrix includes off-diagonal submatrices that represent the covariance between $\mathbf{Z}(t)$ and $\mathbf{X}(t)$. As shown in Preisendorfer (1988, p.161-162), one can then obtain the EOF solution in the conventional manner by diagonalizing the $\mathbf{C}_0^W$ matrix; that is,

$$\mathbf{C}_0^W = \Psi_W \Lambda_W \Psi_W', \quad (90)$$

where the columns of Ψ_W are the eigenvectors (i.e., EOFs) and Λ_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 could be considered simultaneously. However, there is a practical limitation to this procedure since the covariance matrix (89) can easily become very large if the number of observation locations or variables increases. Bretherton et al. (1992) compare

this approach to other multivariate methods such as Canonical Correlation Analysis and Singular Value Decomposition (see Section 5) and find that, in some cases, the multivariate EOF approach has large biases and does not perform well in small signal-to-noise ratio situations.

2.6.3 Extended EOF Analysis

Extended EOFs (e.g., Weare and Nasstrom 1982) are simply multivariate EOFs in which the additional variables are lagged versions of the same process. For example, we could let

$$\mathbf{W}(t) = [\mathbf{Z}(t)' \mathbf{Z}(t-1)]' \quad (91)$$

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

$$\mathbf{C}_1^Z \equiv E[\mathbf{Z}(t) \mathbf{Z}(t-1)']. \quad (92)$$

In this way, we can examine the propagation of EOF spatial patterns in time by noting that the first n eigencoefficients of a particular eigenvector correspond to the time zero representation of that EOF, and the second n eigencoefficients 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. A brief comparison of these three approaches is presented in Section 5.3.

3 Principal Interaction Patterns (PIPs)

Principal Interaction Patterns (PIPs) were originally proposed by Hasselmann (1988) for the continuous time case. He considered a system represented by the state vector $\mathbf{Z}(t)$, whose evolution is governed by a set of first-order equations,

$$\frac{d\mathbf{Z}(t)}{dt} = \mathcal{F}(\mathbf{Z}(t)), \quad (93)$$

where $\mathcal{F}$ is some (possibly) non-linear time-dependent function of $\mathbf{Z}(t)$. The goal is to construct a simplified dynamical model approximating (93) which involves a significantly smaller number

of dimensions $m : m < n$. von Storch et al. (1995) have explained PIPs from the discrete-time perspective using vector-space ideas. We shall adopt the discrete-time approach as well, although from a generalized least-squares point of view.

Our goal is to approximate the dynamical system $\mathbf{Z}(t) \in R^n$ as being driven by a lower-dimensional dynamical system $\mathbf{a}(t) \in R^m$. There is a statistical model underlying PIPs, which we write in the following form

$$\mathbf{Z}(t) = \mathbf{P}\mathbf{a}(t) + \boldsymbol{\epsilon}(t) \quad (94)$$

$$\mathbf{a}(t+1) = \mathcal{F}(\mathbf{a}(t), \boldsymbol{\beta}, t) + \boldsymbol{\eta}(t), \quad (95)$$

where $\mathbf{Z}(t)$ are the observations at locations $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ and at time t ; $\mathbf{P} \equiv [\mathbf{p}_1, \dots, \mathbf{p}_m]$ is an $n \times m$ matrix with column vectors $\mathbf{p}_i$, known as the PIPs; $\mathcal{F}(\cdot)$ denotes a class of models that can be nonlinear in the dynamical variables $\mathbf{a}(t)$ and, additionally, depends on a set of "free" parameters $\boldsymbol{\beta}$; and $\boldsymbol{\epsilon}(t), \boldsymbol{\eta}(t)$ are, in general, unspecified error terms. We shall assume here that $\boldsymbol{\epsilon}(t)$ is multivariate white noise and is uncorrelated with $\boldsymbol{\eta}(t)$.

The PIP analysis is given below as a two-stage procedure.

Stage One:

Define the predictor

$$\hat{\mathbf{Z}}(t) \equiv \mathbf{P}\mathbf{a}(t). \quad (96)$$

Assuming $\mathbf{P}$ known and $\text{var}[\boldsymbol{\epsilon}(t)] \equiv \mathbf{R}$, the generalized least squares estimator of $\mathbf{a}(t)$ can be obtained by minimizing

$$(\mathbf{Z}(t) - \mathbf{P}\mathbf{a}(t))'\mathbf{R}^{-1}(\mathbf{Z}(t) - \mathbf{P}\mathbf{a}(t)) \quad (97)$$

with respect to $\mathbf{a}(t)$. Then the estimator of $\mathbf{a}(t)$ is

$$\hat{\mathbf{a}}(t) = (\mathbf{P}'\mathbf{R}^{-1}\mathbf{P})^{-1}\mathbf{P}'\mathbf{R}^{-1}\mathbf{Z}(t). \quad (98)$$

Stage Two:

Now assume $\mathbf{a}(t)$ is known and seek estimates of $\mathbf{P}$ and $\boldsymbol{\beta}$ by a generalized least squares difference between the derivatives of the $\mathbf{Z}(\cdot)$ process and the predictor (96). Because we are

approaching the problem from a discrete time perspective, we consider the first-differences rather than the derivatives.

$$\mathbf{U}(t) \equiv \mathbf{Z}(t+1) - \mathbf{Z}(t) \quad (99)$$

$$= \mathbf{P}[\mathbf{a}(t+1) - \mathbf{a}(t)] + \boldsymbol{\epsilon}(t+1) - \boldsymbol{\epsilon}(t) \quad (100)$$

$$= \mathbf{P}[\mathcal{F}(\mathbf{a}(t), \boldsymbol{\beta}, t) - \mathbf{a}(t)] + \boldsymbol{\nu}(t), \quad (101)$$

where

$$\boldsymbol{\nu}(t) \equiv \mathbf{P}\boldsymbol{\eta}(t) + \boldsymbol{\epsilon}(t+1) - \boldsymbol{\epsilon}(t). \quad (102)$$

Define

$$\hat{\mathbf{U}} \equiv \mathbf{P}[\mathcal{F}(\mathbf{a}(t), \boldsymbol{\beta}, t) - \mathbf{a}(t)], \quad (103)$$

and

$$\text{var}[\boldsymbol{\nu}(t)] \equiv \mathbf{V}. \quad (104)$$

Then, a generalized least-squares estimator of $\mathbf{P}$ and $\boldsymbol{\beta}$ is obtained by minimizing

$$(\mathbf{U}(t) - \hat{\mathbf{U}}(t))' \mathbf{V}^{-1} (\mathbf{U}(t) - \hat{\mathbf{U}}(t)) \quad (105)$$

with respect to $\mathbf{P}$ and $\boldsymbol{\beta}$.

This minimization is complicated by the fact that $\mathbf{V}$ is a function of $\mathbf{P}$ as can be seen from (102). Furthermore, if $\mathcal{F}(\cdot)$ is nonlinear, then some additional estimation complexity is present. Either way, the minimization of (105) is likely to require iteration. The two-stage formulation suggests that (98) and (105) could be used in an iterative procedure.

3.1 Application of PIPs

Until recently, a full implementation of PIPs had not been performed in the atmospheric sciences. As is usually the case for nonlinear regression with additive errors, the nonlinear function $\mathcal{F}(\cdot)$ should have some physical justification. Realistic models in the atmospheric sciences are quite complex systems of nonlinear equations and this complexity has slowed the implementation of PIPs. However, Achatz et al. (1995) have used PIPs in the examination

of baroclinic wave life cycles. To the best of my knowledge, no one has tried a completely empirical approach to PIPs in which they let the data define the structure of the $\mathcal{F}(\cdot)$ function. One simplification of PIPs, which has been used extensively, is known as Principal Oscillation Patterns (POPs) and will be examined in Section 4.

3.2 Alternate Views of PIPs

We note that equations (94) and (95) could be implemented in a Kalman filter framework. In particular, since the function $\mathcal{F}(\cdot)$ is in general nonlinear, one would need to make use of one of the nonlinear Kalman filter approaches, such as the extended Kalman filter. This technique essentially uses the linear term in the Taylor series expansion of $\mathcal{F}(\cdot)$ (e.g., Grewal and Andrews 1993, p.168-170). Perhaps surprisingly, this approach has not been used in the PIP literature. However, one could argue that this is the approach that has effectively been tried in some of the recent experimental Kalman filter approaches to the data assimilation problem in meteorology (e.g., Miller et al. 1994; Daley 1995), although it has not been recognized as such.

Because the first differences (99) are used in the estimation of parameters, we might also consider developing the model in a multivariate context in terms of $\mathbf{Z}(t)$ and $\mathbf{U}(t)$, where

$$\mathbf{Z}(t) = \mathbf{P}\mathbf{a}(t) + \boldsymbol{\epsilon}(t) \quad (106)$$

$$\mathbf{U}(t) = \mathbf{P}[\mathcal{F}(\mathbf{a}(t), \boldsymbol{\beta}, t) - \mathbf{a}(t)] + \boldsymbol{\nu}(t) \quad (107)$$

$$\mathbf{a}(t+1) = \mathcal{F}(\mathbf{a}(t), \boldsymbol{\beta}, t) + \boldsymbol{\eta}(t), \quad (108)$$

and

$$\boldsymbol{\nu}(t) \equiv \mathbf{P}\boldsymbol{\eta}(t) + \boldsymbol{\epsilon}(t+1) - \boldsymbol{\epsilon}(t). \quad (109)$$

Such a system has not been investigated. However, it seems that it would be superior to the traditional PIP model because it includes cross-covariances between $\mathbf{Z}(t)$ and $\mathbf{U}(t)$, in a manner similar to that for cokriging in the geostatistical literature (e.g., Cressie 1993, p.138-142).

4 Principal Oscillation Patterns (POPs)

Principal Oscillation Patterns (POPs) were originally formulated as a specific case of PIPs by Hasselmann (1988). POPs were reformulated by von Storch et al. (1988) and extended to complex fields by Bürger (1993). Recent years have seen a substantial increase in the number of POP applications in the literature. A comprehensive overview of POPs can be found in von Storch et al. (1995).

In essence, POP analysis assumes that the (multivariate) data field has a temporal autoregressive structure of order one (AR(1)). A chief difference between POPs and other spatio-temporal decompositions (with the exception of PIPs) 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.

4.1 Formulation of POPs

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

$$\mathbf{Z}(t+1) = \mathbf{B}\mathbf{Z}(t) + \boldsymbol{\eta}(t), \quad (110)$$

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 white noise) such that

$$E[\boldsymbol{\eta}(t)\mathbf{Z}(t)'] = \mathbf{0} \quad (111)$$

$$E[\boldsymbol{\eta}(t)] = \mathbf{0} \quad (112)$$

$$E[\boldsymbol{\eta}(t)\boldsymbol{\eta}(\tau)'] \equiv \mathbf{C}^\eta, \quad t = \tau; \mathbf{0}, \text{ otherwise.} \quad (113)$$

Thus, this model is a special case of the PIP model (94) and (95) in which there is no reduction in the order of the dynamical system (i.e., $\mathbf{Z}(t) = \mathbf{a}(t)$) and the two equations collapse into one. The function $\mathcal{F}(\cdot)$ is the matrix product using $\mathbf{B}$, and the system is linear and first-order Markov.

Note that if (110) is post-multiplied by $\mathbf{Z}(t)$ then, upon taking expectations, we obtain

$$E[\mathbf{Z}(t+1)\mathbf{Z}(t)'] = \mathbf{B}E[\mathbf{Z}(t)\mathbf{Z}(t)'] + E[\boldsymbol{\eta}(t)\mathbf{Z}(t)']. \quad (114)$$

Thus, we can solve for $\mathbf{B}$ in

$$\mathbf{B} = \mathbf{E}[\mathbf{Z}(t+1)\mathbf{Z}(t)'](\mathbf{E}[\mathbf{Z}(t)\mathbf{Z}(t)'])^{-1} \quad (115)$$

$$= \mathbf{C}_1^Z[\mathbf{C}_0^Z]^{-1}, \quad (116)$$

where

$$\mathbf{C}_1^Z \equiv \text{cov}[\mathbf{Z}(t+1), \mathbf{Z}(t)] \quad (117)$$

$$\mathbf{C}_0^Z \equiv \text{var}[\mathbf{Z}(t)], \quad (118)$$

and we have assumed $\mathbf{E}[\mathbf{Z}(t)]$ and $\mathbf{E}[\mathbf{Z}(t+1)]$ to be zero, and that $\mathbf{C}_0^Z$ is non-singular.

As shown by Rao (1973, p.43-44), for a non-symmetric matrix $\mathbf{A}_{n \times n}$, the characteristic equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$ has 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

$$\mathbf{A}\mathbf{p}_i = \lambda_i\mathbf{p}_i; \quad i = 1, \dots, n \quad (119)$$

$$\mathbf{A}'\mathbf{q}_i = \lambda_i\mathbf{q}_i; \quad i = 1, \dots, n. \quad (120)$$

Rao then shows that

- $\{\mathbf{p}_i\}$ are linearly independent and so are $\{\mathbf{q}_i\}$,
- $\mathbf{p}_i^*\mathbf{q}_j = 0$ for $i \neq j$, where $*$ denotes the Hermitian transpose (e.g., $\mathbf{a}^*\mathbf{b} = \sum_i a_i \bar{b}_i$),
- 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

$$\mathbf{A} = \mathbf{P}\mathbf{A}\mathbf{D}^{-1}\mathbf{Q}' \quad (121)$$

$$= \sum_{i=1}^n \frac{\lambda_i}{d_i} \mathbf{p}_i \mathbf{q}_i^*, \quad (122)$$

where we have assumed that the latent roots are distinct.

Using (122), we can decompose the AR(1) model matrix $\mathbf{B}$ in (110) as

$$\mathbf{B} = \sum_{i=1}^n \frac{\lambda_i}{d_i} \phi_i \psi_i^*, \quad (123)$$

where $d_i \equiv \psi_i^* \phi_i$, ϕ_i is the right singular vector of $\mathbf{B}$, and ψ_i is the left singular vector of $\mathbf{B}$, corresponding to the latent root $\lambda_i : i = 1, \dots, n$.

Now, note that because

$$\sum_{i=1}^n \frac{\phi_i \psi_i^*}{d_i} = \Phi \mathbf{D}^{-1} \Psi^* \quad (124)$$

$$= \Phi (\Psi^* \Phi)^{-1} \Psi^* \quad (125)$$

$$= \mathbf{I}_{n \times n}, \quad (126)$$

we can write

$$\mathbf{Z}(t) = \sum_{i=1}^n \frac{\phi_i \psi_i^*}{d_i} \mathbf{Z}(t) \quad (127)$$

$$= \sum_{i=1}^n a_i(t) \phi_i, \quad (128)$$

where

$$a_i(t) \equiv \frac{\psi_i^* \mathbf{Z}(t)}{d_i}. \quad (129)$$

Thus, the essence of the POP analysis is the decomposition represented by (128) and (129). The vectors $\{\phi_i\}$ 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. We note that, although the ϕ_i are not orthogonal with themselves, they are orthogonal with the normalized adjoint patterns ψ_i^*/d_i .

4.2 Physical Implication of POPs

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

$$\mathbf{Z}(t+1) = \mathbf{B}\mathbf{Z}(t). \quad (130)$$

The eigenvectors $\phi_i; i = 1, \dots, n$ 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 ϕ_i^* also satisfy the eigen-equation:

$$\mathbf{B}\phi_i^* = \lambda_i^*\phi_i^*. \quad (131)$$

Now, if we multiply (130) by ψ_j^*/d_j , we obtain

$$a_j(t+1) = \frac{\psi_j^*}{d_j} \left[\sum_{i=1}^n \frac{\lambda_i}{d_i} \phi_i \psi_i^* \right] \mathbf{Z}(t) \quad (132)$$

$$= \frac{\psi_j^*}{d_j} \sum_{i=1}^n \lambda_i \phi_i a_i(t) \quad (133)$$

$$= \frac{\psi_j^*}{d_j} \phi_j \lambda_j a_j(t) \quad (134)$$

$$= \lambda_j a_j(t). \quad (135)$$

Equation (135) is a first-order homogeneous difference equation with solution (assuming $a_j(0) = 1$)

$$a_j(t) = (\lambda_j)^t. \quad (136)$$

If λ_j is complex and $i \equiv \sqrt{-1}$, then

$$\lambda_j \equiv \lambda_j^R + i\lambda_j^I, \quad (137)$$

which can be written in polar form as

$$\lambda_j^R = \gamma_j \cos(\omega_j) \quad (138)$$

$$\lambda_j^I = \gamma_j \sin(\omega_j). \quad (139)$$

Then,

$$\lambda_j = \gamma_j e^{i\omega_j}, \quad (140)$$

where

$$\gamma_j = \{(\lambda_j^R)^2 + (\lambda_j^I)^2\}^{1/2}. \quad (141)$$

Thus, (136) and (140) gives

$$a_j(t) = \gamma_j^t e^{i\omega_j t}, \quad (142)$$

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 .

We now decompose the eigenvector (i.e., normal mode) ϕ_j as the sum of a real and an imaginary term:

$$\phi_j = \phi_j^R + i\phi_j^I. \quad (143)$$

Then, noting 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 ϕ_j^R and ϕ_j^I (see e.g., von Storch et al. 1995). That is, the evolution of a damped mode occurs in a succession

$$\dots \rightarrow \phi_j^R \rightarrow -\phi_j^I \rightarrow -\phi_j^R \rightarrow \phi_j^I \rightarrow \phi_j^R \rightarrow \dots \quad (144)$$

with a period of $2\pi/\omega_j$, each stage in (144) occurring a quarter of a cycle apart. Note that 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:

$$\tau_j \equiv -\frac{1}{\ln(\gamma_j)}. \quad (145)$$

4.3 Estimation of POPs

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

In order to perform the POP analysis in practice, the system matrix $\mathbf{B}$ must be estimated. From (116) we see that a method-of-moments (MOM) estimator for $\mathbf{B}$ is

$$\hat{\mathbf{B}} = \hat{\mathbf{C}}_1^Z [\hat{\mathbf{C}}_0^Z]^{-1}, \quad (146)$$

where $\hat{\mathbf{C}}_0^Z$ is a MOM estimator, as shown in (65). Similarly, the (i, j) -th element of $\mathbf{C}_1^Z$ is given by,

$$c_1^Z(\mathbf{s}_i, \mathbf{s}_j) \equiv E(Z(\mathbf{s}_i; t)Z(\mathbf{s}_j; t-1)), \quad (147)$$

so that the MOM estimator is

$$\hat{c}_1^Z(\mathbf{s}_i, \mathbf{s}_j) \equiv \frac{1}{T-1} \sum_{t=2}^T [Z(\mathbf{s}_i; t) - \hat{\mu}_z(\mathbf{s}_i; t)][Z(\mathbf{s}_j; t-1) - \hat{\mu}_z(\mathbf{s}_j; t-1)], \quad (148)$$

where possible choices for the mean estimator are discussed in Section 2.5.

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 we would expect the deterministic normal modes to behave. For example, for damped empirical modes (i.e., $\hat{\gamma}_j \leq 1$, where $\hat{\gamma}_j = |\hat{\lambda}_j|$) we expect the succession (144). 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 $a_j^I(t)$.

4.4 Application of POPs

As with all of the spatio-temporal methods used in the atmospheric sciences, POP analysis can be applied as either a diagnostic or a prognostic tool. Both of these applications will be briefly examined in the following subsections.

4.4.1 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. For example, von Storch et al. (1988) used POPs to consider the equatorial 30-60 day oscillation (Madden and Julian 1971).

4.4.2 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 (135), it is easy to show from (110) that

$$a_j(t+1) = \lambda_j a_j(t) + \tilde{\eta}_j(t), \quad (149)$$

where

$$\tilde{\eta}_j(t) \equiv \frac{\psi_j^* \eta(t)}{d_j}. \quad (150)$$

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 considered 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

$$\hat{a}_j(t+1) = \hat{\gamma}_j e^{i\hat{\omega}_j} \hat{a}_j(t), \quad (151)$$

which then gives

$$\hat{\mathbf{Z}}_j(t+1) \equiv \hat{a}_j(t+1) \phi_j, \quad (152)$$

and hence

$$\hat{\mathbf{Z}}(t+1) = \sum_{j=1}^n \hat{\mathbf{Z}}_j(t+1), \quad (153)$$

or some truncated version of (153). 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 (von Storch et al. 1995). It is not at all apparent that the noise $\{\tilde{\eta}_j(t)\}$ is necessarily “unpredictable”. In fact, (150) shows that under Gaussian white noise assumptions for $\eta(t)$, $\tilde{\eta}_j(t)$ is simply a linear combination of Gaussian random variables, and so must itself be a Gaussian random variable, but *with* spatial dependence. We should then be able to use this dependence to increase our prognostic skill. This is the essence of time series and spatial prediction methodologies.

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. The EOF projection is useful for other reasons as explained below. Examples of POP forecasting can be found in Xu and von Storch (1990) and von Storch and Xu (1990) as well as von Storch et al. (1995). Clearly, the noise reduction referred to above is *ad hoc*; optimal methods should increase the forecasting skill of POPs.

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. Penland (1989) makes 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.

While the POP approach has intuitive appeal in its deterministic form, perhaps the most glaring weakness of the approach is its treatment of error. In general, the error is simply ignored, leading to non-optimal predictions. In an attempt to deal with practical difficulties, a number of *ad hoc* "fix ups" have been used. We then ask the question, can POPs be reformulated in such a way as to account for the presence of error optimally? Kooperberg and O'Sullivan (1994) have recently addressed this issue with what they refer to as predictive oscillation patterns.

4.5 POPs on the EOF Basis

As mentioned in the previous section, before a POP analysis is conducted, the data can be projected onto a truncated set of EOFs in order to reduce the spatial dimension of the system. In that case, it is assumed that noise is then excluded from the analysis (although not optimally as we have shown in Section 2.4). The EOF expansion also improves the estimation from a practical numerical perspective, analogous to the use of principal components in linear regression (e.g., Draper and Smith 1981, p. 327-331). In the POP case, the estimate of $\mathbf{B}$ given in (146) contains the inverse of $\mathbf{C}_0^Z$, which could be very unstable if the dimension of n is large and the data are noisy. In that case, the smallest (and presumably, physically uninteresting) eigenvalues and associated eigenvectors of $\mathbf{C}_0^Z$ dominate the decomposition of $\mathbf{B}$ (Kooperberg and O'Sullivan 1994). Thus, by projecting the data onto the first K EOFs, we can reduce the spatial dimension from n to K and obtain a diagonal $\mathbf{C}_0^Z$ matrix that is quite easy to invert.

4.6 Extensions of POPs

In this section, some extensions of POPs are considered. Some, such as continuous time POPs, Complex POPs, and Cyclostationary POPs, have been considered in the literature. Others, such as POPs in the presence of measurement error, nonstationary POPs, and two-field POPs have not been considered, to the best of my knowledge.

4.6.1 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 (e.g., Penland 1989). Penland (1989) and collaborators (Penland and Ghil 1993; Penland and Magorian 1993) have taken this approach and demonstrated its usefulness, particularly with regard to prognostic applications.

4.6.2 Complex POPs

Complex POP (CPOP) analysis was introduced by Bürger (1993) as an extension of conventional POP analysis. Bürger notes that while POP analysis is able to model *traveling oscillations*, they are unable to model *standing oscillations* (see Section 2.6.1 for an explanation of traveling and standing oscillations). In fact, he shows the inherent impossibility of modeling standing oscillations in first-order linear systems. So 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.

Just as for the CEOF analysis (Section 2.6.1), we write the system in terms of a new state process

$$\mathbf{W}(t+1) = \mathbf{G}\mathbf{W}(t) + \boldsymbol{\eta}(t), \quad (154)$$

where

$$\mathbf{W}(t) \equiv \mathbf{Z}(t) + i\tilde{\mathbf{Z}}(t), \quad (155)$$

and $\tilde{\mathbf{Z}}(t)$ is the Hilbert transform (Section 2.6.1) of $\mathbf{Z}(t)$. The CPOP analysis then proceeds in a similar manner to the POP analysis. However, since $\mathbf{G}$ is complex, its eigenvectors (i.e., the CPOPs) do not appear in complex conjugate pairs. von Storch et al. (1995) provides an excellent review of this approach.

4.6.3 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. An excellent discussion of the analysis of periodically correlated atmospheric time series can be found in Lund et al. (1995).

Blumenthal (1991) first published the idea of using periodically correlated statistics (i.e., cyclostationarity) in the POP framework. His approach is summarized in von Storch et al. (1995). In this case, the cyclostationary process is written as

$$\mathbf{Z}(t, \tau + 1) = \mathbf{A}(\tau)\mathbf{Z}(t, \tau) + \tilde{\boldsymbol{\eta}}(t, \tau), \quad (156)$$

where t and τ are integers such that t counts the cycles (e.g., years) and $\tau = 1, \dots, m$ counts the position within each cycle (e.g., months). Then, $\mathbf{Z}(t, \tau + m) \equiv \mathbf{Z}(t + 1, \tau)$ and $\mathbf{A}(\tau + m) \equiv \mathbf{A}(\tau)$. Thus we can write the cyclostationary POP model for $\tau = 1, \dots, m$ as

$$\mathbf{Z}(t + 1, \tau) = \mathbf{B}(\tau)\mathbf{Z}(t, \tau) + \boldsymbol{\eta}(t, \tau), \quad (157)$$

where

$$\mathbf{B}(\tau) \equiv \mathbf{A}(\tau + m - 1)\mathbf{A}(\tau + m - 2) \dots \mathbf{A}(\tau), \quad (158)$$

and

$$\boldsymbol{\eta}(t, \tau) \equiv \sum_{j=0}^{m-1} \tilde{\boldsymbol{\eta}}(t, \tau + j). \quad (159)$$

We then obtain a set of POPs (i.e., eigenvectors) for each τ .

4.6.4 POPs in the Presence of Measurement Error

The issue of measurement error in the model leading to POPs has not been addressed in the literature. If we assume that the true dynamical system of interest is given by the signal $\mathbf{Y}(t)$ in the presence of additive measurement error, then

$$\mathbf{Z}(t) = \mathbf{Y}(t) + \boldsymbol{\epsilon}(t). \quad (160)$$

Thus, we can write the POP model in a state-space formulation

$$\mathbf{Z}(t) = \mathbf{Y}(t) + \boldsymbol{\epsilon}(t) \quad (161)$$

$$\mathbf{Y}(t+1) = \boldsymbol{\beta}\mathbf{Y}(t) + \boldsymbol{\eta}(t). \quad (162)$$

This formulation has not been considered from a POP perspective but, because measurement error is so pervasive, it is an important area to pursue. It naturally would lead to a Kalman-filter implementation. Perhaps, if interested in a particular phenomenon, one could include one's prior belief about the process by specifying $\boldsymbol{\beta}$ in a Bayesian setting. This would lead to a hierarchical Bayesian analysis of the system (161),(162).

4.6.5 Two-Field POPs

We might also assume that there is a linear relationship between the process $\mathbf{Z}(t)$ and another process (of possibly different dimension) at a previous time $\mathbf{X}(t-1)$ given by:

$$\mathbf{Z}(t+1) = \boldsymbol{\beta}\mathbf{X}(t) + \boldsymbol{\nu}(t), \quad (163)$$

where $\boldsymbol{\beta}$ is the two-field system matrix, and $\boldsymbol{\nu}(t)$ is some error process, independent of $\mathbf{X}(t)$. This model is essentially a two-field extension of POPs. We would then be interested in the decomposition of the matrix $\boldsymbol{\beta}$, which can be written as

$$\boldsymbol{\beta} = \mathbf{C}_1^{Z,X} [\mathbf{C}_0^X]^{-1}, \quad (164)$$

where

$$\mathbf{C}_1^{Z,X} \equiv E[\mathbf{Z}(t+1)\mathbf{X}(t)'] \quad (165)$$

$$\mathbf{C}_0^X \equiv E[\mathbf{X}(t)\mathbf{X}(t)']. \quad (166)$$

Note that this model does not include any dynamic structure on $\mathbf{X}(\cdot)$.

Neither the physical implications nor the statistical aspects of this model have been investigated. In particular, a comparison between this approach and the two-field spatio-temporal canonical correlation analysis (see below) would be interesting to pursue.

4.6.6 Non-Stationary POPs

As discussed previously, the stationary (in time) assumption in POP analysis forces damped oscillatory solutions, which present a problem for realistic prognostic applications. This suggests that one should investigate non-stationary vector AR models and their implications concerning growing modes (i.e., modes with increasing amplitudes) in POP-type analyses.

5 Space-Time Canonical Correlation Analysis (CCA)

Canonical Correlation Analysis (CCA) is a long-standing multivariate statistical technique (Hotelling 1936) 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 (e.g., Glahn 1968; Nicholls 1987; Barnett and Preisendorfer 1987), in the forecast of El Niño (Graham and Michaelsen 1987; Barnston and Ropelewski 1992), and the forecast of long-range temperature and precipitation (Barnston 1994). Bretherton et al. (1992) performed an intercomparison of methods for finding coupled patterns in climate data (including CCA and multivariate EOFs) and were supportive of a variant of CCA known (unfortunately) as Singular Value Decomposition (SVD). This is related to, but not the same as, the well-known matrix algebra technique of the same name. Some authors (e.g., Cherry 1994) have pointed out that there are difficulties with the interpretation of CCA and SVD results, particularly the tendency for the methods to produce spurious spatial patterns.

In the following, we shall review the traditional two-field space-time CCA approach used in the atmospheric sciences, briefly describe its estimation, examine the Singular Value Decomposition relative of CCA, consider the one-field time-lagged CCA approach, which is related to the

minimum/maximum autocorrelation factor (MAF) technique in statistics (Shapiro and Switzer 1989) and, finally, consider two extensions of CCA that have not been examined previously, CCA in the presence of measurement error and CCA in continuous space.

5.1 Two-Field Spatial-Temporal CCA

We assume that in addition to the process $Z(\mathbf{s};t)$ we are given another related process $X(\mathbf{s};t)$ with a possibly different spatial domain, but the same temporal domain. We further assume discrete space and time and zero means:

$$E[\mathbf{Z}(t)] = \mathbf{0}_{n \times 1} \quad (167)$$

$$E[\mathbf{X}(t)] = \mathbf{0}_{m \times 1}, \quad (168)$$

where $\mathbf{X}(t) \equiv (X(\mathbf{r}_1;t), \dots, X(\mathbf{r};t))'$ and $\mathbf{Z}(t)$ is defined as before in (11). Furthermore, we define the covariances

$$\mathbf{C}_0^Z \equiv E[\mathbf{Z}(t)\mathbf{Z}(t)']_{n \times n} \quad (169)$$

$$\mathbf{C}_0^X \equiv E[\mathbf{X}(t)\mathbf{X}(t)']_{m \times m} \quad (170)$$

$$\mathbf{C}_0^{Z,X} \equiv E[\mathbf{Z}(t)\mathbf{X}(t)']_{n \times m}, \quad (171)$$

which are invariant in time. We then define linear combinations of each data field

$$a_k(t) = \phi_k' \mathbf{Z}(t) \quad (172)$$

$$b_k(t) = \psi_k' \mathbf{X}(t), \quad (173)$$

where $k = 1, \dots, \min\{m, n\}$. Now, define the k -th *canonical correlation* as

$$r_k \equiv \text{corr}[\phi_k' \mathbf{Z}(t), \psi_k' \mathbf{X}(t)] \quad (174)$$

$$= \frac{\psi_k' \mathbf{C}_0^{Z,X} \psi_k}{(\phi_k' \mathbf{C}_0^Z \phi_k)^{1/2} (\psi_k' \mathbf{C}_0^X \psi_k)^{1/2}}. \quad (175)$$

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 (175) and have unit variance. The second pair of canonical variables are then the linear combinations $a_2(t)$ and $b_2(t)$ that are uncorrelated

with $a_1(t)$ and $b_1(t)$, have unit variance, and maximize (175). 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 (175).

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

$$\mathbf{C}_0^Z = (\mathbf{C}_0^Z)^{1/2}(\mathbf{C}_0^Z)^{1/2} \quad (176)$$

$$\mathbf{C}_0^X = (\mathbf{C}_0^X)^{1/2}(\mathbf{C}_0^X)^{1/2}. \quad (177)$$

Then, we can write

$$r_k^2 = \frac{[\tilde{\phi}_1' (\mathbf{C}_0^Z)^{-1/2} \mathbf{C}_0^{Z,X} (\mathbf{C}_0^X)^{-1/2} \tilde{\psi}_1]^2}{(\tilde{\phi}_1' \tilde{\phi}_1)(\tilde{\psi}_1' \tilde{\psi}_1)}, \quad (178)$$

where

$$\tilde{\phi}_1 \equiv (\mathbf{C}_0^Z)^{1/2} \phi_1 \quad (179)$$

$$\tilde{\psi}_1 \equiv (\mathbf{C}_0^X)^{1/2} \psi_1 \quad (180)$$

are normalized weights. The problem is now reduced to finding $\tilde{\phi}_1$ and $\tilde{\psi}_1$ that maximize (178). 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 (e.g., Johnson and Wichern 1992, p. 463) that r_1^2 is the largest singular value of

$$(\mathbf{C}_0^Z)^{-1/2} \mathbf{C}_0^{Z,X} (\mathbf{C}_0^X)^{-1/2}, \quad (181)$$

and $\tilde{\phi}_1$ and $\tilde{\psi}_1$ are the left and right singular vectors corresponding to r_1^2 , respectively (see Section 4.1). Then, we can write

$$\phi_1 \equiv (\mathbf{C}_0^Z)^{-1/2} \tilde{\phi}_1 \quad (182)$$

$$\psi_1 \equiv (\mathbf{C}_0^X)^{-1/2} \tilde{\psi}_1. \quad (183)$$

We can also obtain the time series of canonical variables (for $t = 1, \dots, T$):

$$a_1(t) = \phi_1' \mathbf{Z}(t) \quad (184)$$

$$b_1(t) = \psi_1' \mathbf{X}(t). \quad (185)$$

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 of (181). Then, ϕ_k and ψ_k can be obtained analogous to (182) and (183), as well as $a_k(t)$ and $b_k(t)$ analogous to (184) and (185).

We can also examine the correlation between the canonical variable time series and the original data. We obtain

$$r[a_k(t), \mathbf{Z}(t)] \equiv \text{corr}[a_k(t), \mathbf{Z}(t)] \quad (186)$$

$$= \phi_k' \mathbf{C}_0^Z [\text{diag}(\mathbf{C}_0^Z)]^{-1/2} \quad (187)$$

$$r[b_k(t), \mathbf{X}(t)] \equiv \text{corr}[b_k(t), \mathbf{X}(t)] \quad (188)$$

$$= \psi_k' \mathbf{C}_0^X [\text{diag}(\mathbf{C}_0^X)]^{-1/2}, \quad (189)$$

where $[\text{diag}(\mathbf{C}_0^Z)]^{-1/2}$ is an $n \times n$ diagonal matrix with (i, i) -th element given by $\{\text{var}[Z(s_i; t)]\}^{-1/2}$, 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 (187) and (189) 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, we can define the k -th left and right *heterogeneous correlation maps*, respectively, as

$$r[a_k(t), \mathbf{X}(t)] \equiv \text{corr}[a_k(t), \mathbf{X}(t)] \quad (190)$$

$$= \phi_k' \mathbf{C}_0^{X,Z} [\text{diag}(\mathbf{C}_0^X)]^{-1/2} \quad (191)$$

$$r[b_k(t), \mathbf{Z}(t)] \equiv \text{corr}[b_k(t), \mathbf{Z}(t)] \quad (192)$$

$$= \psi_k' \mathbf{C}_0^{Z,X} [\text{diag}(\mathbf{C}_0^Z)]^{-1/2}. \quad (193)$$

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

5.2 Estimation of CCA

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

$$\hat{\mathbf{C}}_Z \equiv \frac{1}{T} \sum_{t=1}^T \mathbf{Z}(t) \mathbf{Z}(t)' \quad (194)$$

$$\hat{\mathbf{C}}_X \equiv \frac{1}{T} \sum_{t=1}^T \mathbf{X}(t) \mathbf{X}(t)' \quad (195)$$

$$\hat{\mathbf{C}}_{Z,X} \equiv \frac{1}{T} \sum_{t=1}^T \mathbf{Z}(t) \mathbf{X}(t)', \quad (196)$$

where both the $Z(\cdot)$ and $X(\cdot)$ processes are assumed to have zero means. Possible estimates of the mean are described in Section 2.5. Then, the estimated singular values and singular vectors are obtained from the numerical singular value decomposition of

$$(\hat{\mathbf{C}}_0^Z)^{-1/2} \hat{\mathbf{C}}_0^{Z,X} (\hat{\mathbf{C}}_0^X)^{-1/2}. \quad (197)$$

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 (see Section 4.4.2). As in that case, we get 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 stability. 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 (e.g., Preisendorfer 1988).

5.3 Modifications of CCA

This section describes several modifications to CCA. In particular, there is a brief description of the popular Singular Value Decomposition relative of CCA, a description of the time-lagged version of CCA and its relationship to MAFs in statistics, as well as a brief exam-

ination of the previously unexamined issues of CCA in the presence of measurement error and CCA in continuous space.

5.3.1 Singular Value Decomposition Relative of CCA

Bretherton et al. (1992) popularized a close relative of CCA which they unfortunately call Singular Value Decomposition, and which we will refer to as SVD/CCA. In essence, SVD/CCA consists of finding $\check{\phi}_k$ and $\check{\psi}_k$ such that if

$$\check{a}_k(t) = \check{\phi}_k' \mathbf{Z}(t) \quad (198)$$

$$\check{b}_k(t) = \check{\psi}_k' \mathbf{X}(t), \quad (199)$$

then

$$c_k \equiv \text{cov}[\check{a}_k(t), \check{b}_k(t)] \quad (200)$$

$$= \check{\phi}_k' \mathbf{C}_0^{Z,X} \check{\psi}_k \quad (201)$$

is maximized given uncorrelatedness with the $(k-1)$ previous values as well as the additional constraints

$$\check{\phi}_k' \check{\phi}_k = \check{\psi}_k' \check{\psi}_k = 1; \quad k = 1, \dots, n. \quad (202)$$

Clearly, it is these constraints that account for the difference between SVD/CCA and CCA. Cherry (1994) discusses some of the implications of these additional constraints.

Then, the problem can be restated as one of finding the $\check{\phi}_k$ and $\check{\psi}_k$ that are, respectively, the right and left singular vectors from the singular value decomposition of $\mathbf{C}_0^{Z,X}$, where c_k is the k -th singular value.

5.3.2 Time-Lagged CCA

When only one process (say $\mathbf{Z}(\cdot)$) 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 could then be useful for prognostic applications. In this case, we would like the vectors ϕ_k and

ψ_k such that the marginal correlation between

$$a_k(t) = \phi'_k \mathbf{Z}(t), \quad (203)$$

and

$$b_k(t + \tau) = \psi'_k \mathbf{Z}(t + \tau) \quad (204)$$

is maximized. To do this we obtain the k -th left and right singular vectors from the singular value decomposition applied to

$$(\mathbf{C}_0^Z)^{-1/2} \mathbf{C}_\tau^Z (\mathbf{C}_0^Z)^{-1/2}, \quad (205)$$

where

$$\mathbf{C}_\tau^Z \equiv E[\mathbf{Z}(t)\mathbf{Z}(t + \tau)'], \quad (206)$$

and where we have assumed $\mathbf{Z}(\cdot)$ has zero mean. Analogous to (182) and (183), 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 (Section 4) and to the minimum/maximum autocorrelation factor (MAF) approach in statistics (Shapiro and Switzer 1989; Cressie and Helderbrand 1994). In the MAF case, one is interested in the eigenvectors proportional to those obtained from the singular value decomposition of

$$(\mathbf{C}_0^Z)^{-1} \mathbf{V}, \quad (207)$$

where the matrix $\mathbf{V}$ is the first-difference correlation matrix:

$$\mathbf{V} \equiv E[(\mathbf{Z}(t) - \mathbf{Z}(t + 1))(\mathbf{Z}(t) - \mathbf{Z}(t + 1))'] \quad (208)$$

$$= 2\mathbf{C}_0^Z - \mathbf{C}_1^Z - \mathbf{C}_1^{Z'} \quad (209)$$

$$= 2\mathbf{C}_0^Z \left[\mathbf{I} - (\mathbf{C}_0^Z)^{-1} \left(\frac{\mathbf{C}_1^Z + (\mathbf{C}_1^Z)'}{2} \right) \right], \quad (210)$$

where we have assumed temporal invariance in obtaining (210). 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, we note that we must perform the singular value decomposition on the following matrices, respectively:

$$\mathbf{C}_1^Z (\mathbf{C}_0^Z)^{-1} \quad (211)$$

$$(\mathbf{C}_0^Z)^{-1/2} \mathbf{C}_1^Z (\mathbf{C}_0^Z)^{-1/2} \quad (212)$$

$$(\mathbf{C}_0^Z)^{-1} \left[\frac{\mathbf{C}_1^Z + (\mathbf{C}_1^Z)'}{2} \right]. \quad (213)$$

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 (211) and (212) and that the singular vectors from the CCA decomposition (212) are equivalent to those from the POP decomposition (211) scaled by the matrix $(\mathbf{C}_0^Z)^{-1/2}$. The relationship between the eigenvalues and eigenvectors of the MAF matrix (213) 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 (see Shapiro and Switzer 1989), a symmetric matrix can be obtained in (213) 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., $\psi_k = \phi_k; k = 1, \dots, n$), so we are able to consider the linear combinations $\phi_k' \mathbf{Z}(t)$ and $\phi_k' \mathbf{Z}(t + \tau)$, which is a useful feature. Additional investigation of the similarities and differences of these approaches would be beneficial.

5.4 CCA in the Presence of Measurement Error

Although the effect of sampling variability in CCA is well-known in the statistics literature (e.g., Johnson and Wichern 1993, Section 10.4), to the best of my knowledge, the issue of measurement error in CCA has not been considered. Assume that $\mathbf{Z}(t)$ and $\mathbf{X}(t)$ can be

written as,

$$\mathbf{Z}(t) = \mathbf{Y}(t) + \boldsymbol{\epsilon}(t) \quad (214)$$

$$\mathbf{X}(t) = \mathbf{W}(t) + \boldsymbol{\gamma}(t), \quad (215)$$

where $\mathbf{Y}(t)$ and $\mathbf{W}(t)$ are the true physical processes, and $\boldsymbol{\epsilon}$ and $\boldsymbol{\gamma}$ are additive measurement error. Then the following covariance function relationships are obtained:

$$\mathbf{C}_0^Z = \mathbf{C}_0^Y + \mathbf{C}_0^\epsilon \quad (216)$$

$$\mathbf{C}_0^X = \mathbf{C}_0^W + \mathbf{C}_0^\gamma, \quad (217)$$

where we have used an assumption that the errors are uncorrelated with their respective physical processes. Furthermore, if we assume that the two error processes are uncorrelated with each other, then we obtain the relationship,

$$\mathbf{C}_0^{Y,W} = \mathbf{C}_0^{Z,X}. \quad (218)$$

Now, let

$$\tilde{a}_k(t) \equiv \tilde{\phi}'_k \mathbf{Y}(t) \quad (219)$$

$$\tilde{b}_k(t) \equiv \tilde{\psi}'_k \mathbf{W}(t). \quad (220)$$

Following the same methodology as before, it is straightforward to show that the maximization of the squared correlation between (219) and (220) is accomplished through the singular value decomposition of

$$(\mathbf{C}_0^Z - \mathbf{C}_0^\epsilon)^{-1/2} \mathbf{C}_0^{Z,X} (\mathbf{C}_0^X - \mathbf{C}_0^\gamma)^{-1/2}. \quad (221)$$

Then we may ask two questions:

- How does the singular value decomposition of (221) compare to the singular value decomposition of (181) (i.e., without measurement error)?
- How do we estimate the measurement error covariances in (221)?

The first question may be answered approximately by expanding $(\mathbf{C}_0^Z - \mathbf{C}_0^\epsilon)^{-1/2}$ and $(\mathbf{C}_0^X - \mathbf{C}_0^\gamma)^{-1/2}$ in a power series, and truncating. The second question can be answered if separate data are available for assessment of the measurement error. Alternatively, one could fit a parametric model to the covariances, as is commonly done in geostatistical studies (e.g., Cressie 1993, Section 2.4).

5.5 CCA in Continuous Space

To the best of my knowledge, no one has considered CCA in a continuous space framework, analogous to the Karhunen-Loève (K-L) EOF approach. Of course, from a physical perspective, most of the spatial processes we deal with in atmospheric science are continuous, with observations at certain locations within a given domain. From this point of view, most of the data are missing and those that are available are collected together in a multivariate vector of observations to which multivariate methods are applied.

Consider the continuous spatial, discrete temporal processes:

$$Z(\mathbf{s}; t) : \mathbf{s} \in D_Z, t \in \{1, \dots, T\} \quad (222)$$

$$X(\mathbf{r}; t) : \mathbf{r} \in D_X, t \in \{1, \dots, T\}. \quad (223)$$

We then define

$$a_k(t) \equiv \int_{D_Z} \phi_k(\mathbf{s}) Z(\mathbf{s}; t) d\mathbf{s} \quad (224)$$

$$b_k(t) \equiv \int_{D_X} \psi_k(\mathbf{r}) X(\mathbf{r}; t) d\mathbf{r}. \quad (225)$$

We look for $\phi_k(\cdot)$ and $\psi_k(\cdot)$ such that $\text{corr}[a_k(t), b_k(t)]$ is maximized, subject to the usual CCA orthogonality conditions. Let

$$r_k \equiv \text{corr}[a_k(t), b_k(t)] \quad (226)$$

$$= \text{corr} \left[\int_{D_Z} \phi_k(\mathbf{s}) Z(\mathbf{s}; t) d\mathbf{s}, \int_{D_X} \psi_k(\mathbf{r}) X(\mathbf{r}; t) d\mathbf{r} \right] \quad (227)$$

$$= \frac{\int_{D_X} \int_{D_Z} \phi_k(\mathbf{s}) \psi_k(\mathbf{r}) c_0^{Z,X}(\mathbf{s}, \mathbf{r}) d\mathbf{s} d\mathbf{r}}{\left[\int_{D_Z} \int_{D_Z} \phi_k(\mathbf{s}) \phi_k(\mathbf{s}) c_0^Z(\mathbf{s}, \mathbf{s}) d\mathbf{s} d\mathbf{s} \right]^{1/2} \left[\int_{D_X} \int_{D_X} \psi_k(\mathbf{r}) \psi_k(\mathbf{r}) c_0^X(\mathbf{r}, \mathbf{r}) d\mathbf{r} d\mathbf{r} \right]^{1/2}}, \quad (228)$$

where

$$c_0^Z(\mathbf{s}, \mathbf{s}) \equiv E[Z(\mathbf{s}; t)Z(\mathbf{s}; t)] \quad (229)$$

$$c_0^X(\mathbf{r}, \mathbf{r}) \equiv E[X(\mathbf{s}; t)X(\mathbf{s}; t)] \quad (230)$$

$$c_0^{Z,X}(\mathbf{s}, \mathbf{r}) \equiv E[Z(\mathbf{s}; t)X(\mathbf{r}; t)], \quad (231)$$

and we have assumed temporal invariance.

Intuitively, we would expect that there is a form of K-L singular value decomposition formulation that could be applied to (228) to aid in the maximization. Further effort is necessary to demonstrate this.

6 Conclusion

In Sections 2-5 we have given the fundamentals of the EOF, PIP, POP, and CCA techniques as applied in the atmospheric sciences. We have seen that there are many physical and statistical considerations that must be made when applying these techniques to data. In particular, we must decide if the problem should be considered discrete or continuous, with measurement error or without, prognostic or diagnostic, or involve spatial prediction or smoothing. We also must consider what is being optimized and whether there should be a dynamical (i.e., temporally dependent) component, what estimation strategy is appropriate, what are the effects of sampling variability, and whether Bayesian (e.g., Kalman filters) ideas should be considered. Some of these issues have been considered while others have only been touched upon. In fact, it is clear that most of these issues should be considered simultaneously. However, in practice this is rarely done.

Scientists are typically interested in their own scientific problems and not in what they may perceive as arcane statistical issues. They generally want simple, fast, yet powerful tools to help them achieve their scientific goals. Unfortunately, methods that are easy to use are usually very limited in their ability to deal with complicated problems. The spatio-temporal methods outlined in this review provide some very powerful diagnostic and prognostic tools for dealing

with high-dimensional spatio-temporal data sets. Although these methods are generally easy to implement in their traditional forms, in many cases extra effort is required to ensure that they provide the optimal results for the problem at hand. It is hoped that this review has identified some of the areas where additional efforts should be focused.

References

- Achatz, U., G. Schmitz, and K.-M. Greisiger, 1995: Principal interaction patterns in baroclinic wave life cycles. *J. Atmos. Sci.*, **52**, 3201-3213.
- Barnett, T.P., 1977: The principal time and space scales of the Pacific trade wind fields. *J. Atmos. Sci.*, **34**, 221-236.
- Barnett, T.P., 1983: Interaction of the monsoon and Pacific trade wind system at interannual time scales. Part I: The equatorial zone. *Mon. Wea. Rev.*, **111**, 756-773.
- Barnett, T.P., and R.W. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.*, **115**, 1825-1850.
- Barnston, A.G., 1994: Linear statistical short-term climate predictive skill in the Northern hemisphere. *J. Climate*, **7**, 1513-1564.
- Barnston, A.G., and C.F. Ropelewski, 1992: Prediction of ENSO episodes using canonical correlation analysis. *J. Climate*, **5**, 1316-1345.
- Blumenthal, B., 1991: Predictability of a coupled ocean-atmosphere model. *J. Climate*, **4**, 766-784.
- Braud, I., and Ch. Obled, 1991: On the use of empirical orthogonal function (EOF) analysis in the simulation of random fields. *Stochastic Hydrol. Hydraul.*, **5**, 125-134.
- Bretherton, C.S., C. Smith, and J.M. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate data. *J. Climate*, **5**, 541-560.
- Buell, C.E., 1972: Integral equation representation for factor analysis. *J. Atmos. Sci.*, **28**, 1502-1505.
- Buell, C.E., 1975: The topography of empirical orthogonal functions. *Preprints Fourth Conf. Prob. Stats. Atmos. Sci.*, American Meteorological Society, 188-193.
- Bürger, G., 1993: Complex principal oscillation pattern analysis. *J. Climate*, **6**, 1972-1986.

- Chatfield, C., 1989: *The Analysis of Time Series: An Introduction*, Fourth Edition, Chapman and Hall, 241pp.
- Cherry, S., 1994: Some comments on singular value decomposition and canonical correlation analysis. *J. Climate*, (submitted for review).
- Christenson, W.I., Jr., and R.A. Bryson, 1966: An investigation of the potential of component analysis for weather classification. *Mon. Wea. Rev.*, **94**, 697-709.
- Cohen, A. and R.H. Jones, 1969: Regression on a random field. *J. of the Amer. Stat. Assoc.*, **64**, 1172-1182.
- Cressie, N.A.C., 1993: *Statistics for Spatial Data, Revised Edition*, Wiley, 900pp.
- Cressie, N. and J.D. Helerbrand, 1994: Multivariate spatial statistical models. *Geographical Systems*, **1**, 179-188.
- Daley, R., 1995: Estimating the wind field from chemical constituent observations: Experiments with a one-dimensional extended Kalman filter. *Mon. Wea. Rev.*, **123**, 181-198.
- Davis, R.E., 1976: Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific ocean. *J. Physical Oceanography*, **6**, 249-266.
- Draper, N.R., and H. Smith, 1981: *Applied Regression Analysis*, Second Edition, Wiley, 709pp.
- Freiberger, W. and U. Grenander, 1965: On the formulation of statistical meteorology. *Review of the International Statistical Institute*, **33**, 59-86.
- Glahn, H.R., 1968: Canonical correlation and its relationship to discriminant analysis and multiple regression. *J. Atmos. Sci.*, **25**, 23-31.
- Graham, N.E., and J. Michaelson, 1987: An investigation of the El Nino-Southern Oscillation cycle with statistical models. 1. Predictor field characteristics. *J. Geophys. Res.*, **92**, 14251-14270.
- Gray, M., 1981: On the stability of temperature eigenvector patterns. *J. Climate*, **1**, 273-281.
- Grewal, M.S., and A.P. Andrews, 1993: *Kalman Filtering: Theory and Practice*, Prentice-Hall, 381pp.
- Hasselmann, K., 1988: PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.*, **93**, 11015 - 11021.
- Horel, J.D., 1984: Complex principal component analysis: theory and examples. *J. Clim. Appl. Meteor.*, **23**, 1660-1673.

- Hotelling, H., 1933: Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, **24**, 417-441, 498-520.
- Hotelling, H., 1936: Relations between two sets of variables. *Biometrika*, **28**, 321-377.
- Johnson, R.A., and D.W. Wichern, 1992: *Applied Multivariate Statistical Analysis*, Prentice-Hall, 641 pp.
- Karl, T.R., A.J. Koscielny, and H.F. Diaz, 1982: Potential errors in the application of principal component (eigenvector) analysis to geophysical data. *J. Appl. Meteor.*, **21**, 1183-1186.
- Kooperberg, C. and F. O'Sullivan, 1994: Predictive Oscillation Patterns: a synthesis of methods for spatial-temporal decomposition of random fields. *Technical Report No. 277, Department of Statistics, University of Washington, Seattle.*
- Kutzbach, J.E., 1967: Empirical eigenvectors of sea level pressure, surface temperature and precipitation complexes over North America. *J. Appl. Meteor.*, **6**, 791-802.
- Lawley, D.N., 1956: Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, **43**, 128-136.
- Loève, M., 1963: *Probability Theory*, Van Nostrand Company, 685 pp.
- Lorenz, E.N., 1956: Empirical orthogonal functions and statistical weather prediction. *Sci. Rept. No. 1, Statistical Forecasting Project, MIT*, 49 pp.
- Lund, R.B., H.L. Hurd, P. Bloomfield, and R. Smith, 1995: Climatological time series with periodic correlation. *J. Climate*, **8**, 2787-2809.
- Madden, R.A., and P.R. Julian, 1971: Detection of a 40-50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28**, 702-708.
- Miller, R.N., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.*, **51**, 1037-1056.
- Nicholls, N. 1987: The use of canonical correlation to study teleconnections. *Mon. Wea. Rev.*, **115**, 393-399.
- North, G.R., T.L. Bell, R.F. Cahalan, and F.J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.*, **110**, 699-706.
- Obled, Ch., and J.D. Creutin, 1986: Some developments in the use of empirical orthogonal functions for mapping meteorological fields. *J. Clim. Appl. Meteor.*, **25**, 1189-1204.
- Papoulis, A., 1965: *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Inc., pp. 457-461.

- Peagle, J.N., and R. B. Haslam, 1982: Statistical prediction of 500mb height field using eigenvectors. *J. Appl. Meteor.*, **21**, 127-138.
- Penland, C., 1989: Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Wea. Rev.*, **117**, 2165-2185.
- Penland, C., and M. Ghil, 1993: Forecasting Northern Hemisphere 700-mb geopotential height anomalies using empirical normal modes. *Mon. Wea. Rev.*, **121**, 2355-2372.
- Penland, C., and T. Magorian, 1993: Prediction of Niño 3 sea surface temperatures using linear inverse modeling. *J. Climate*, **6**, 1067-1076.
- Preisendorfer, R.W., 1988: *Principal Component Analysis in Meteorology and Oceanography*, Elsevier, 425 pp.
- Preisendorfer, R.W., and T.P. Barnett, 1983: Numerical model-reality intercomparison test using small-sample statistics. *J. Atmos. Sci.*, **40**, 1884-1896.
- Rao, C.R., 1973: *Linear Statistical Inference and its Applications*, Wiley.
- Rasmusson, E.M., P.A. Arkin, and W.Y. Chen, 1981: Biennial variation in surface temperature over the United States as revealed by singular decomposition. *Mon. Wea. Rev.*, **109**, 587-598.
- Richman, M.B., 1981: Obliquely rotated principal components: An improved meteorological map typing technique? *J. Appl. Meteor.*, **20**, 1145-1159.
- Richman, M.B., 1986: Review article: rotation of principal components. *J. Climatol.*, **6**, 293-335.
- Shapiro, D.E., and P. Switzer, 1989: Extracting time trends from multiple monitoring sites. *Technical Report No. 132*, Department of Statistics, Stanford University, Stanford, CA.
- Trenberth, K.E., and W.-T. K. Shin, 1984: Quasi-biennial fluctuations in sea level pressures over the Northern Hemisphere. *Mon. Wea. Rev.*, **112**, 761-777.
- von Storch, H., G. Bürger, R. Schnur, and J.-S. von Storch, 1995: Principal oscillation patterns: a review. *J. Climate*, **8**, 377-400.
- von Storch, H., and G. Hannoshock, 1985: Statistical aspects of estimated principal vectors (EOFs) based on small sample sizes. *J. Clim. Appl. Meteor.*, **24**, 716-724.
- von Storch, H., T. Bruns, I. Fischer-Bruns, and K.H. Hasselmann, 1988: Principal Oscillation Pattern analysis of the 30 to 60 day oscillation in a GCM. *J. Geophys. Res.*, **93**, 11022-11036.

- von Storch, H., and J.-S. Xu, 1990: Principal oscillation pattern analysis of the tropical 30- to 60-day oscillation. Part I Definition of an index and its prediction. *Climate Dynamics*, **4**, 175-190.
- Wallace, J.M., 1972: Empirical orthogonal representation of time series in the frequency domain. Part II: Application to the study of tropical wave disturbances. *J. Appl. Meteor.*, **11**, 893-900.
- Wallace, J.M., and R.E. Dickinson, 1972: Empirical orthogonal representation of time series in the frequency domain. Part I: Theoretical considerations. *J. Appl. Meteor.*, **11**, 887-892.
- Weare, B.C., and J.S. Nasstrom, 1982: Examples of extended empirical orthogonal function analysis. *Mon. Wea. Rev.*, **110**, 481-485.
- Xu, J.-S., 1992: On the relationship between the stratospheric QBO and the tropospheric SO. *J. Atmos. Sci.*, **49**, 725-734.
- Xu, J.-S., and H. von Storch, 1990: Predicting the state of the Southern Oscillation using principal oscillation pattern analysis. *J. Climate*, **3**, 1316-1329.

GENERAL CONCLUSION

Meteorological and climatological processes typically show variability over both space and time. Consequently, scientists who study these processes require spatio-temporal models that can characterize this variability. Such characterizations then help the scientists to make inference about the nature of the processes, and eventually aid in their prediction. Thus, my goal in this dissertation has been to help bridge the gap between the need for advanced spatio-temporal statistical methods in the atmospheric sciences, and the development of such methods in the statistical sciences.

The dissertation consists of a background chapter followed by three additional chapters, each of which represents an independent paper. The background chapter is an overview of spatio-temporal statistical methods commonly used in the atmospheric sciences, from a statistical point of view. The first paper uses some simple harmonic analysis ideas to make spatial inference about a possible physical mechanism for the observed semiannual oscillation in the Northern Hemisphere extratropical height field. The second paper uses some advanced cyclostationary time series techniques to explore the seasonal variability of lower stratospheric mixed Rossby-gravity waves over the tropical Pacific. Finally, the third paper presents a new spatio-temporal statistical model that can be used to predict in both space and time. The results from each of these studies are described briefly below.

In the first chapter, a comprehensive overview of spatio-temporal methods that have received attention in the atmospheric science literature is presented. In particular, the focus is on the Empirical Orthogonal Function (EOF), Principal Interaction Pattern (PIP), Principal Oscillation Pattern (POP), and spatio-temporal Canonical Correlation Analysis (CCA) methods.

Particular attention is given to the physical and statistical considerations that must be made when applying these techniques to data. Specifically, these include the consideration of discrete or continuous space, the inclusion of measurement error, whether the application is prognostic or diagnostic, and whether the goal is spatial prediction, temporal prediction, or smoothing. We also consider what it is that is being optimized in a particular method, whether a dynamical component is appropriate, and whether Bayesian (i.e., Kalman filter) ideas should be considered. In general, most of these issues must be considered simultaneously. Unfortunately, this is rarely done in practice. Throughout the paper, we present several potential questions that are deserving of additional research. These generally are concerned with continuous space applications, and with the inclusion of measurement error in the analysis.

In the second chapter (i.e., the first paper) simple harmonic analysis is used to make diagnostic inference about the spatial variation of the semiannual component of the atmospheric general circulation. In particular, based on an examination of the spatial distribution of the maximum semiannual oscillation (SAO) amplitudes in the Northern Hemisphere (NH) extratropical 500-hPa height field, we concluded that this oscillation exhibits a very dominant zonally asymmetric east-west spatial structure. This led us to conclude that the stationary eddies inherent in the NH general circulation might be a useful tool with which to explain the SAO. Indeed, a comparison between the stationary eddies and the SAO showed that the NH midlatitude SAO can be explained almost entirely by the spatial and temporal asymmetries in the annual variation of the stationary eddies. It was suggested that the mechanism for the SAO in the NH extratropics is simply a result of land-sea contrasts, similar to the well-known explanation of the Southern Hemisphere (SH) SAO. However, the NH SAO is likely due to east-west contrasts between the continental land masses and oceans in the NH, while the SH SAO is due to the north-south contrast between the Antarctica land mass and the surrounding ocean. Thus, we have unified the conceptual view of the atmospheric SAO in both hemispheres. It was suggested that additional research should be focused on modeling studies with which these hypotheses could be verified, as well as additional focus on the NH polar SAO.

The third chapter (i.e., the second paper) is concerned with the seasonal variability of mixed Rossby-gravity waves (MRGWs) in the lower stratosphere over the tropical western Pacific. The study of these waves is important because they are believed to be a critical forcing mechanism of the quasi-biennial oscillation (QBO) in the tropical stratosphere. Recent observational studies have suggested that MRGWs generally do not show semiannual variability, but rather have a single seasonal peak, depending on geographic location. However, since these waves are believed to be associated with tropical convective activity, and since this convective activity exhibits semiannual peaks, we hypothesized that MRGWs should show semiannual peaks. Thus, we employed some relatively sophisticated time series analysis techniques to study the seasonal variability of these waves. These analyses used long time records of wind data in the lower stratosphere at four tropical Pacific observation stations. Specifically, seasonally varying cross-spectral analysis suggested that there are significant twice-yearly peaks in the v -wind power and the mean squared coherence between the u - and v -winds, with peaks occurring in the winter-early spring and in summer-early fall. In addition, the seasonally varying phase associated with the mean squared coherence analysis suggested that there is convergence of horizontal momentum flux associated with these waves, and that the sign of the convergence is opposite during the two seasonal maxima. This convergence of momentum flux in MRGWs has not been identified previously and deserves additional research effort. Furthermore, an autoregressive cyclic spectral estimate showed that the frequency of the maximum v -wind power in the MRGW frequency band shifts seasonally. This shift may be related to the seasonal variation of MRGWs, although further effort is required to prove such a claim.

In the fourth chapter (i.e., the third paper), a new spatio-temporal statistical model is proposed that attempts to consider the influence of both temporal and spatial variability. This model is mainly concerned with prediction, unlike the traditional spatio-temporal methods used in the atmospheric science literature (and outlined in the first chapter of this dissertation), which were primarily designed for diagnostic applications. The model is developed within the framework of continuous space and discrete time. The model then assumes a first-order

Markov temporal dynamic structure in conjunction with a spatially descriptive colored noise process. With the inclusion of a measurement error equation, we developed a spatio-temporal Kalman filter prediction algorithm that allowed us to predict in time and at spatial locations at which we do not have observations. The model prediction equation is quite general and includes a simple kriging analog as a special case. The model was shown to predict well with a simulated spatio-temporal data set, and was shown to be superior to simple kriging applied independently at each time. The model was then applied to a large precipitation data set in the South China Sea region. Predictions of precipitation over the data-sparse South China Sea seemed to capture the dynamic variation of the spatial precipitation field. Since this work is quite new, there are many possible avenues for future research. These were outlined in the paper.

ACKNOWLEDGEMENTS

I would never have been able to accomplish this task without the understanding and support of my family. First and foremost, I would like to express my thanks to Carolyn, the value of whose love and support throughout the past five years has been incalculable. I am forever grateful for her sacrifice as we placed much of our lives on hold through this period so that I could pursue a dream. I also thank my two wonderful children, Olivia and Nathan, who have never known a different life. I have appreciated their understanding and patience throughout this process and for their always being there to brighten my day.

I am eternally grateful to my parents, Bayliss and Irene Wikle, for instilling in me the value of education, and for giving me the freedom and support to search for my own path. Their support and love has shown me what it means to be a parent. In addition, I would like to thank my in-laws for their support, encouragement, and understanding over the years. I am exceedingly fortunate to have *two* such wonderful families.

I wish to thank by co-major professors, T.-C. (Mike) Chen and Noel Cressie. In addition to being top scientists in their respective fields, both have mastered the art of teaching, and the more elusive skill of mentoring. In countless informal conversations, Dr. Chen has taught me more than just the art of diagnostic analysis in atmospheric science, he has taught me the *philosophy of science* and the importance of knowing the history of your field. His support and encouragement in good times and in bad have been a source of strength. Dr. Cressie has provided countless points of wisdom and encouragement over the last couple of years, as well. In addition to opening my eyes to the wonders of spatial analysis, he has taught me how to think like a statistician and has provided a model of professionalism and scientific optimism

that I shall strive to emulate throughout my career. I especially appreciate the extra time, effort, and encouragement he has given over the last weeks during the completion of my degree.

For taking the time out of their busy schedules, I would also like to thank my committee members: Dr. Roland Madden of the National Center for Atmospheric Research, Dr. Peter Sherman, Dr. Raymond Arritt, and Dr. William Gutowski. In particular, Dr. Sherman has spent a great deal of time teaching me about modern digital signal processing, and provided encouragement throughout the growing pains of my first scientific paper. I also would like to thank Dr. Madden for generously giving his time to teach me how to analyze physical time series and for setting the standard for using such analyses to answer questions about the physical world.

Of course, much of what has made the last five years such a positive experience was the friendship and support of my fellow students. In particular, I would like to thank Jean Pelkey, Alkhalil Adoum, Craig Clark, Zaitao Pan, Susan Kiehne, and John Iselin.

Finally, all of this would not have been possible without the generous financial support of the Graduate Fellowships for Global Change Program of the U.S. Department of Energy, as administered by Oak Ridge Institute for Science and Education. Rather than just talk about the need for interdisciplinary training in the sciences, they have funded it! I hope that I can prove that their substantial investment has not been in vain.