Preface

These notes are the result of courses I have taught at the University of Maryland in 1982 and at Texas A&M University in 1993 and 1995. They draw heavily on material I hastily read usually the night before the lectures in 1982. At the time I was desperately trying to learn the subject matter better for myself, since my own research required that I do so. I relied heavily on books by Bulmer, Jenkins and Watts, Blackman and Tukey, and an Econometrics book by Maddala. In this first edition of the notes, I tried mostly to just get the equations typed. The notes need considerable fleshing out before a book could be produced. In addition there are likely to be a good many errors, so the notes should be used with caution. Just because the notes are printed with the help of \TeX{}does not mean they are correct, nor does it mean that all the references are in the correct spots (authors please forgive!). Of course, I am grateful to the students who had to endure my red-pencil correcting the drafts as I lectured. GRN, Bryan, Texas, July 22, 2004
## Contents

1 Elementary Prob & Stats Review .......................................................... 9  
  1.1 Moment Generating Function ......................................................... 10  
  1.1.1 Binomial Distribution ............................................................ 11  
  1.1.2 Poisson Distribution ............................................................. 11  
  1.2 Normal Distribution ................................................................. 12  
  1.2.1 Sum of Two Normal Variates .................................................. 13  
  1.3 Central Limit Theorem ............................................................... 13  
  1.3.1 Exponential Distribution ...................................................... 15  
  1.4 $\chi^2$ Distribution ................................................................. 18  
  1.5 Hypothesis Testing ................................................................. 18  
  1.5.1 Example ............................................................................ 20  
  1.6 Confidence Intervals ............................................................... 21  
  1.7 Student $t$ Distribution ............................................................ 22  
  1.8 $F$ Distribution ................................................................. 23  

2 Regression ................................................................................. 27  
  2.1 Curve Fitting ................................................................. 27  
  2.1.1 Statistical Model ............................................................ 28  
  2.1.2 Confidence Intervals for $a, b$ .................................................. 29  
  2.1.3 Example of Simple Regression ............................................... 30  
  2.2 Apportioning Variance ............................................................ 32  
  2.3 Correlation ........................................................................ 33  
  2.4 Significance Test for $r$ ........................................................... 39  
  2.4.1 When $\rho \neq 0$ ............................................................ 40  
  2.4.2 Correlation between two curves ............................................. 41  
  2.4.3 Max Likelihood and Linear Regression ................................... 41  
  2.5 Multiple Linear Regression ....................................................... 42  
  2.6 Multivariate Normal Distribution ............................................. 45  
  2.6.1 Confidence regions ........................................................... 45  
  2.6.2 2 Dimensional Example ......................................................... 45
B Note on Statistical Extrapolation of Fields 157
  B.1 Optimal Filter Formulation .......................... 157
  B.2 Examples (incomplete) ............................. 160
    B.2.1 One Dimensional Interpolation .................. 160
    B.2.2 Prediction ................................. 161

C Principal Oscillation Patterns 163
  C.1 Adjoint Patterns ................................. 164
  C.2 Stochastic Model ............................... 164
  C.3 Estimating the System Matrix .................... 165
  C.4 Cyclostationary POPs ............................ 165

D Singular Value Decomposition 167
  D.1 Covariance ...................................... 167
Chapter 1

Elementary Prob & Stats Review

This chapter is not much more than a list of information that is useful to atmospheric scientists. There are many excellent books that cover the subject. Here I assume everyone is familiar with the elements of probability theory. But not everyone is familiar with the term *random variable*. We consider a variable \( X \) to be a random variable if it takes on different values depending upon some procedure involving an element of uncertainty such as a drawing. Most of the random variables (rv’s) we will deal with are continuous which means they can take on a continuous set of values over some interval of real numbers. The *cumulative distribution* \( F(x) \) is the probability that \( X < x \).

\[
F(x) = \text{Prob}(X < x) \quad (1.1)
\]

In general the cumulative distribution is a monotonically increasing function of \( x \), approaching unity at the upper end of the range of \( x \) [pic of a typical \( F(x) \)]. The probability density function \( f(x) \) is the probability of \( X \) lying in an infinitesimal interval centered at \( x \) per unit interval.

\[
\int_{a}^{b} f(x) \, dx = \text{Prob} \ a \leq X \leq b \quad (1.2)
\]

The *expected value* of a rv is

\[
E(X) = \langle X \rangle = \int x f(x) \, dx = \mu_x \quad (1.3)
\]
The expected value of a function of a random variable \( g(X) \) is similarly defined:

\[
\langle g(X) \rangle = \int g(x)f(x) \, dx \quad (1.4)
\]

The variance is an important measure of the dispersion of a distribution:

\[
\text{Var}(X) = \langle (X - \mu_x)^2 \rangle = \int (x - \mu_x)^2 f(x) \, dx \quad (1.5)
\]

The variance is the second centered moment of the distribution, the mean being the first. In general

\[
\mu_n = \langle (X - \mu_1)^n \rangle \quad (1.7)
\]

Uncentered moments are defined by

\[
\mu'_n = \langle X^n \rangle \quad (1.8)
\]

There are many important distributions encountered in atmospheric science. In what follows we list several of them. But first we define the moment generating function since it plays such a key role.

### 1.1 Moment Generating Function

It is defined by

\[
M(t) = \langle e^{tx} \rangle = \int f(x)e^{tx} \, dx = \sum_{n=0}^{\infty} \mu_n \frac{t^n}{n!} \quad (1.9)
\]

The last comes from Taylor expanding \( e^{tx} \) and integrating term by term. The moment generating function is particularly useful in studying sums of independent random variables. For example, consider

\[
Z = X + Y \quad (1.12)
\]

where \( X \) and \( Y \) are two independent rv’s. That is, the joint distribution factors

\[
f(x, y) = f_1(x)f_2(y) \quad (1.13)
\]
1.1. MOMENT GENERATING FUNCTION

Then the moment generating function is

\[ M_Z(t) = \langle e^{t(X+Y)} \rangle \quad (1.14) \]

\[ = \langle e^{tX} \rangle \cdot \langle e^{tY} \rangle \quad (1.15) \]

\[ = M_x(t)M_y(t) \quad (1.16) \]

Hence, if we know the moment generating functions for \( X \) and \( Y \) separately, we can find same for \( Z \). This will turn out to be important in a number of applications later.

1.1.1 Binomial Distribution

The binomial distribution is encountered when we consider a system with two possible outcomes: \( S \) and \( F \)

\[ P = \text{Prob of Success}, \quad Q = 1 - P \quad (1.17) \]

We consider \( n \) trials of the elementary process. In this set of \( n \) trials let there be \( X \) successes. Then \( X \) is a discrete random variable; in fact it is an integer taking on the possible values 0, \ldots, \( n \).

\[ P(x) = \frac{n!}{x!(n-x)!} P^x Q^{n-x} \quad (1.18) \]

\[ E(X) = nP \quad (1.19) \]

\[ \text{Var}(X) = nPQ \quad (1.20) \]

Proportion of successes:

\[ Y = \frac{X}{n} \quad (1.21) \]

\[ E(Y) = P \quad (1.22) \]

\[ \text{Var}(Y) = PQ/n \quad (1.23) \]

1.1.2 Poisson Distribution

We frequently encounter a situation where there are two outcomes and one of them is very rare. The appropriate limit is the binomial distribution where \( P \) very small and \( n \) is large, such that \( \mu = nP \) is finite.

\[ P(x) = \frac{e^{-\mu} \mu^x}{x!} \quad (1.24) \]

\[ M(t) = e^{-\mu e^t} \quad (1.25) \]

\[ E(X) = \mu \quad (1.26) \]

\[ \text{Var}(X) = \mu \quad (1.27) \]
1.2 Normal Distribution

This distribution is so commonly encountered that one should memorize its properties.

\[ f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}, \quad -\infty < x < \infty \] (1.28)

The distribution depends on only two parameters, \( \mu \) and \( \sigma \). It can be shown that

\[ \mu_x = \mu \] (1.29)
\[ \text{Var}(X) = \sigma^2 \] (1.30)
\[ M(t) = e^{\mu t + \frac{1}{2} \sigma^2 t^2} \] (1.31)

It can then be established that the even centered moments are

\[ \mu_r = 1 \cdot 3 \cdot \ldots \cdot (r-1)\sigma^r \] (1.32)

The odd centered moments vanish. For a normally distributed variable with mean \( \mu \) and variance \( \sigma^2 \), we say \( X \sim N(\mu, \sigma^2) \).

Because of the form of the moment generating function, it is easy to show that linear combinations of independent gaussian variables are also gaussian.
1.2.1 Sum of Two Normal Variates

Consider the sum of two normally distributed random variables:

\[ Y = X_1 + X_2 \]  

(1.33)

with \( \mu_1, \mu_2, \sigma_1, \sigma_2 \). The moment generating function for \( Y \) is

\[
M_Y(t) = e^{\mu_1 t + \frac{1}{2} \sigma_1^2 t^2} \cdot e^{\mu_2 t + \frac{1}{2} \sigma_2^2 t^2} \\
= e^{(\mu_1 + \mu_2) t + \frac{1}{2} (\sigma_1^2 + \sigma_2^2) t^2}
\]  

(1.34) \hspace{1cm} (1.35)

It follows that

\[ Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \]  

(1.36)

It is easy to generalize this to a sum of \( n \) gaussian variates.

1.3 Central Limit Theorem

Consider the sum of a large number of independent, identically distributed (iid) rv's

\[ Y = \sum_{n} X_n \]  

(1.37)
We clearly have

\[ \mu_Y = \sum_{n=1}^{N} \mu_n \quad (1.38) \]

\[ \text{Var}(Y) = \sum_{n=1}^{N} \text{Var}(X_n) \quad (1.39) \]

or

\[ \sigma_Y^2 = \sum_{n} \sigma_n^2 = N \sigma^2 \quad (1.40) \]

where \( \sigma^2 \equiv \sigma_n^2 \). This last since the variates are identically distributed. For simplicity in what follows let us subtract off the mean of \( Y \) so that the new variable (same name, of course) has zero mean. At the same time the means of the individual contributors will also be subtracted out so that each has zero mean. If we choose to use a standardized variable (one whose variance is unity), we could define

\[ Z \equiv \frac{Y}{\sigma_Y} = \frac{Y}{\sigma \sqrt{N}} \quad (1.42) \]

Now consider the Moment Generating Functions for \( Y \) and \( Z \). We can easily show that

\[ M_Z(t) = M_Y \left( \frac{t}{\sigma \sqrt{N}} \right) \quad (1.43) \]

The individual moment generating functions multiply together to give

\[ M_Y(t) = \prod_{n=1}^{N} M_n(t) = (M_n(t))^N \quad (1.44) \]

By taking the ln we get

\[ \ln M_Y(t) = \sum_{n=1}^{N} \ln M_n(t) = N \ln M_n(t) \quad (1.45) \]

Considering \( Z \) instead of \( Y \):

\[ \ln M_Z(t) = N \ln M_n \left( \frac{t}{\sigma \sqrt{N}} \right) \quad (1.46) \]

For finite values of \( t \) and large values of \( \sigma \sqrt{N} \):

\[ M_n \left( \frac{t}{\sigma \sqrt{N}} \right) \approx 1 + \frac{\sigma^2 t^2}{2\sigma^2 N} + O \left( \frac{1}{N^2} \right) + \cdots \quad (1.47) \]
1.3. CENTRAL LIMIT THEOREM

and since

\[ \ln(1 + \epsilon) \approx \epsilon, \; \epsilon << 1 \] \hspace{1cm} (1.48)

We can write

\[ \ln M_Z(t) \approx \sum_n \frac{\sigma_n^2 t^2}{2\sigma_Y^2} + \cdots \] \hspace{1cm} (1.49)

\[ \approx \frac{t^2}{2} + \cdots \] \hspace{1cm} (1.50)

The terms which have been dropped are of higher order in \( \frac{1}{N} \). This means that \( Z \) is approximately distributed as

\[ f(z) \approx \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \] \hspace{1cm} (1.51)

The approximation gets better as \( N \) gets larger. This is only a weak version of the CLT. The variables do not have to be identically distributed. It is important, however, that the variances all exist. Of course, in some important cases this will not happen. Consider for example the Cauchy distribution

\[ f(x) = \frac{1}{\pi} \frac{b}{b^2 + x^2} \] \hspace{1cm} (1.52)

which has a zero mean, but infinite variance.

1.3.1 Exponential Distribution

This distribution occurs in examining the waiting times between events that are governed by a Poisson process.

\[ f(x) = \frac{1}{b} e^{-x/b} \] \hspace{1cm} (1.53)

\[ F(x) = 1 - e^{-x/b} \] \hspace{1cm} (1.54)

\[ M(t) = \frac{1}{1 - bt} \] \hspace{1cm} (1.55)

\[ E(X) = b \] \hspace{1cm} (1.56)

\[ \text{var}(X) = b^2 \] \hspace{1cm} (1.57)
Figure 1.3: Probability density function for the Cauchy distribution with $b = 1$.

Figure 1.4: Cumulative density function for the Cauchy distribution with $b = 1$. 
1.3. CENTRAL LIMIT THEOREM

Figure 1.5: Probability density function for the exponential distribution with $b = 1$.

Figure 1.6: Cumulative density function for the exponential distribution with $b = 1$. 
1.4 χ² Distribution

Consider next the random variable which is the sum of squares of gaussian standard variates:

\[ Y = \sum_{n=1}^{\nu} Z_n^2, \quad Z_n \sim N(0, 1) \quad (1.58) \]

The variable \( Y \) is said to be a \( \chi^2[\nu] \) variate. We can work from the \( \chi^2[1] \) case since the means and variances will add for the \( \nu \) terms. It turns out that

\[ \mu_1 = 1 \quad (1.59) \]
\[ \mu_2 = 2 \quad (1.60) \]

Then for the \( \chi^2[\nu] \) we have

\[ \mu_1^{[\nu]} = \nu \quad (1.61) \]
\[ \mu_2^{[\nu]} = 2\nu \quad (1.62) \]

The probability density function turns out to be

\[ f(y) = \frac{y^{\frac{\nu}{2}-1}e^{-y/2}}{A(\nu)} \quad (1.63) \]

where

\[ A(\nu) = 1 \times 3 \times 5 \ldots (\nu - 1) \times \sqrt{2\pi}, \quad \nu \text{ odd} \quad (1.64) \]
\[ = 2 \times 4 \times (\nu - 2) \times 2 = 2^{\frac{\nu}{2}} \frac{1}{2} \nu! \quad (1.65) \]

or

\[ A(\nu) = 2^{\nu/2} \Gamma(\nu/2) \quad (1.66) \]

where

\[ \text{Gamma}(t) \equiv \int_{0}^{\infty} y^{t-1}e^{-y} \, dy \quad (1.67) \]

is the Gamma function. Note that \( \Gamma(n) = (n-1)! \), for \( n \) a positive integer.

1.5 Hypothesis Testing

Often we want to know if a series of realizations of a variate come from a certain distribution. For example, suppose we want to know whether the variate \( X \) comes from \( N(0, 1) \). We are at liberty to evaluate \( N \) realizations
1.5. HYPOTHESIS TESTING

Figure 1.7: Probability density function for the $\chi^2$ distribution with $\nu = 2$.

Figure 1.8: Probability density function for the $\chi^2$ distribution with $\nu = 4$. 
of \( X : x_1, x_2, \ldots, x_N \). The arithmetic average of \( X \), \( \bar{X} \) should be distributed as \( N(0, \frac{1}{N}) \). Hence if we take the arithmetic average of our data and find that it lies outside the interval \( \pm \frac{2}{\sqrt{N}} \), it would be extremely improbable that the \( N \) realizations were drawn from \( N(0, 1) \), since the area under the normal distribution outside this interval is only about 5%. This concept is the basis for hypothesis testing. One proceeds by assuming the Null Hypothesis: that the realizations do not come from \( N(0, 1) \). If the mean lies outside some preassigned interval, we agree beforehand to reject the null hypothesis. Of course, there is the possibility that the null hypothesis is true if the test statistic \( \bar{X} \) did lie outside the chosen range. Often one chooses the interval such that only 5% or 1% of the area lies in the rejection region.

1.5.1 Example

Consider a climate change experiment. After making a very long run of a climate model (the control run), we establish the global mean and its standard deviation. We form the standardized variate, \( Z \), which has the mean subtracted out and is normalized by the standard deviation of the
1.6 CONFIDENCE INTERVALS

When we are trying to estimate the value of some parameter pertaining to the underlying distribution, such as $\mu$ or $\sigma^2$, we often do not want to use the formalism of hypothesis testing. Rather it is more important to
establish an interval in which the value of the estimated parameter is to lie with some probability. For example in estimating \( \mu \) for a distribution with an arithmetic mean, \( \frac{1}{N} \sum x_i \), we have a standard error \( \sigma_s = \frac{s}{\sqrt{N}} \). The 95% confidence interval for the estimated value of \( \mu \) is \( \pm \frac{s}{\sqrt{N}} \). This means that with many realizations of averages based upon \( N \) drawings we can expect that 95% will lie within the confidence interval.

### 1.7 Student \( t \) Distribution

One weakness in the hypothesis testing procedure outlined above is that it assumes that we know the variance of the underlying distribution which in the above example was unity. In practice we may only know the underlying variance approximately based on the data we have. An unbiased estimate of the standard deviation for \( n \) realizations is

\[
 s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \\
 \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i
\]  

(1.68)  

(1.69)

We could ask how the statistic

\[
 U \equiv \frac{\bar{x}}{s}
\]  

(1.70)

is distributed for a given \( n \), assuming the original distribution from which the \( x_i \) were drawn was normal. Fortunately, this problem has been solved and will be discussed in the next section.

First, however note that

\[
 (n - 1)s = \sum_{i=1}^{n} (x_i - \bar{x})^2
\]  

(1.71)

is a \( \chi^2_{n-1} \) variate. The number of degrees of freedom is \( n - 1 \) instead of \( n \) because of the presence of \( \bar{x} \). By expanding and examining the sum it can be shown that this term makes the sum equivalent to a sum of only \( n - 1 \) squares. We say one degree of freedom has been removed by this constraint.

Here we consider a variate

\[
 T = \frac{Z}{\sqrt{Y/\nu}}
\]  

(1.72)
Figure 1.11: Probability density function for the $t_{\nu}$ distribution with $\nu = 5$.

where $Z \sim N(0,1)$ and $Y \sim \chi^2_{\nu}$. It can be shown that

$$f(t) = \frac{A(\nu + 1)}{A(\nu)\sqrt{2\pi\nu}} \times \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

(1.73)

which is called Student’s $t$ distribution.

Returning now to the problem of the last section, we can take $Z$ to be the average from the $n$ readings divided by the standard deviation: $\sqrt{n\bar{x}}$. The denominator for this problem is $(n-1)s/(n-1) = s$. Hence, the $T$ statistic for this problem is

$$T = \frac{\sqrt{n\bar{x}}}{s}$$

(1.74)

Since $s \rightarrow \sigma$ for $n \rightarrow \infty$ we can assert that the $t$ distribution should approach the normal distribution when the sample size is large enough. The figures illustrate this for some large values of $n$.

1.8 F Distribution

Another very convenient test is on the sample variances of two populations. We can collect two variances $s_1$ and $s_2$ and see if they come from the same
pdf. Since \( s_1 \) and \( s_2 \) are \( \chi^2 \) variates with \( n_1 - 1 \) and \( n_2 - 1 \) degrees of freedom, we can form the ratio and have a variate with a convenient distribution.

The so-called F variate is the ratio of two \( \chi^2 \) variates normalized by their respective dof’s:

\[
V = \frac{Y_1/\nu_1}{Y_2/\nu_2}
\]  

(1.75)
References


Exercises: Elements

Uniform Distribution:

\[
f(x) = \begin{cases} 
0, & x \leq -\frac{1}{2} \\
1, & -\frac{1}{2} < x < \frac{1}{2} \\
0, & x \geq \frac{1}{2}
\end{cases}
\]

1. Consider the uniform distribution above. Show that a) \( \mu = 0 \), b) \( \sigma^2 = \frac{1}{12} \).

2. Show that the Moment Generating Function, \( M(t) = \left( e^{t/2} - e^{-t/2} \right)/t \).

3. Consider the variate \( Y = (U_1 + U_2 + \cdots + U_N)/N \), where the \( U_n \) are uniformly distributed as above. Find the mean and variance of \( Y \).

4. By expanding the exponentials in \( M_Y(t) \) and recollecting terms, show how the distribution of \( Y \) approaches the normal distribution \( N(0, \frac{1}{12N}) \) as \( N \) becomes large.

5. Relate the problem above to estimating the mean of a uniformly distributed variate \( U \) with \( N \) pieces of data.
Chapter 2

Regression

2.1 Curve Fitting

Suppose we are given data in pairs of the form \(x_1, Y_1; x_2, Y_2; \ldots; x_N, Y_N\), where we have used lower case to indicate that the \(x_i\) are fixed numbers and the \(Y_i\) are random variables. It is natural and perhaps useful to ask about ‘fitting’ a straight line through the data

\[ y = a + bx \]  

(2.1)

(called the regression line). What is the best way to choose the coefficients \(a, b\)? These free parameters are to be adjusted to provide the best fit in some sense. One way to proceed is to look at the difference between each \(Y_i\) and the ‘predicted’ value \(a + bx_i\) and compute the sum of squares of ‘errors’ (SSE)

\[ SSE = e^2 = \sum_{n=1}^{N} (Y_i - a - bx_i)^2 \]  

(2.2)

The individual errors are the lengths of vertical line segments joining the points on a scattergram to the regression line. We can adjust \(a, b\) to make \(e^2\) as small as possible. The equations determining the minimum of the surface \(e^2(a, b)\) are \(\partial e^2 / \partial a = 0, \partial e^2 / \partial b = 0\):

\[ \sum (Y_i - a - bx_i) = 0 \]  

(2.3)

\[ \sum x_i (Y_i - a - bx_i) = 0 \]  

(2.4)

or

\[ \bar{y} - a - b \bar{x} = 0 \]  

(2.5)
\[ xy - a\bar{x} - bx^2 = 0 \] (2.6)

or finally:

\[ a = \bar{y} - b\bar{x} \] (2.7)

\[ b = \frac{\bar{xy} - \bar{x}\bar{y}}{x^2 - \bar{x}^2} \] then (2.8)

\[ a = \frac{\bar{y}x^2 - xy\bar{x}}{x^2 - \bar{x}^2} \] (2.10)

The last can also be written

\[ b = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} \] (2.11)

\[ = \frac{S_{xy}}{S_{xx}} \] (2.12)

2.1.1 Statistical Model

It is very important from here to keep in mind that we have tacitly assumed a model for this process

\[ Y_i = \alpha + \beta x_i + E_i \] (2.13)

where \( Y_i \) and \( E_i \) are rv’s while the \( x_i \) are not rv’s. The \( Y_i \) have a certain pdf and it depends upon the value of \( x_i : f(Y|x) \), where the last denotes the conditional pdf for \( Y \) given the value of \( x \). We usually think of the \( E_i \sim N(0, \sigma^2) \) and \( E_i E_j = \sigma^2 \delta_{ij} \). The line \( y = \alpha + \beta x \) is just the expected value of \( Y \) given \( x \). For large enough \( N \) we expect \( a \rightarrow \alpha, b \rightarrow \beta \).

In fact \( a \) and \( b \) are the estimators of \( \alpha \) and \( \beta \). From one realization or sample (a set of realizations of the pairs \( Y_i, x_i \)) to another we may think of \( A \) and \( B \) as rv’s which are estimators of \( \alpha \) and \( \beta \). We may rearrange to obtain

\[ B = \frac{\sum(x_i - \bar{x})(Y_i - \bar{Y})}{\sum(x_i - \bar{x})^2} \] (2.14)

\[ = \frac{\sum(x_i - \bar{x})Y_i}{\sum(x_i - \bar{x})^2} \] (2.15)

\[ = \sum_{i=1}^{N} \gamma_i Y_i \] (2.16)
2.1. CURVE FITTING

a linear combination of the \( Y_i \). It follows that \( B \) is normally distributed. Note again that the \( x_i \) are not rv's but are just fixed coefficients in the last expression. We can find \( E(B) \):

\[
\mu_B = \frac{\sum (x_i - \bar{x})E(Y_i)}{\sum (x_i - \bar{x})^2}
\]

(2.17)

\[
= \frac{\sum (x_i - \bar{x})(\alpha + \beta x_i)}{\sum (x_i - \bar{x})^2}
\]

(2.18)

\[
= \beta
\]

(2.19)

and variance

\[
\sigma_B^2 = \frac{\sum (x_i - \bar{x})^2 \sigma_Y^2}{[\sum (x_i - \bar{x})^2]^2}
\]

(2.20)

\[
= \frac{\sum (x_i - \bar{x})^2}{S_{xx} \sigma^2}
\]

(2.21)

\[
= \frac{\sigma^2}{S_{xx}}
\]

(2.22)

Note that \( S_{xx} \approx N \sigma_x^2 \); hence, it is a good estimator. Similarly, \( \mu_A = \alpha \) and

\[
\sigma_A^2 = \frac{\sum x_i^2}{N \sum (x_i - \bar{x})^2} \sigma^2
\]

(2.23)

\[
= \frac{\sigma^2}{S_{xx} \sigma^2}
\]

(2.24)

where

\[
V(E_i) = \sigma^2
\]

(2.25)

the variance of the random error variable. Note that both \( \sigma_A^2 \propto 1/N \) and \( \sigma_B^2 \propto 1/N \).

2.1.2 Confidence Intervals for \( a, b \)

We know that \( A, B \) are normally distributed so we could proceed to test significance, etc., if we only knew \( \sigma^2 \). At best we have an estimate of it through \( SSE \). Recall that

\[
SSE = \sum_{n=1}^{n} (y_i - a - bx_i)^2 = \sum_{n=1}^{n} E_i^2
\]

(2.26)
It turns out that $SSE/(n - 2)$ is the best unbiased estimator of $\sigma^2$. Define

$$s^2 = \frac{SSE}{n - 2} \quad (2.27)$$

Now $(n - 2)s^2/\sigma^2$ is a $\chi^2$ variate with $n - 2$ dof. (We must subtract 2 since both $a$ and $b$ were fitted from the data).

Now look at the rv $B$.

$$B - \beta \sim N \left(0, \frac{\sigma^2}{S_{xx}}\right) \quad (2.28)$$

Hence,

$$(B - \beta)\frac{\sqrt{S_{xx}}}{\sigma} \sim N(0, 1) \quad (2.29)$$

At last we can divide this $N(0, 1)$ variate by $s/\sigma$ to obtain a Student $t$ variate

$$\frac{(B - \beta)\sqrt{S_{xx}}}{s} = t_{[n-2]} \quad (2.30)$$

where $s = \sqrt{\frac{SSE}{n - 2}} \quad (2.31)$

The confidence interval $(1 - \alpha) \cdot 100\%$ for $\beta$ is then

$$b - \frac{t_{\alpha/2}s}{\sqrt{S_{xx}}} < \beta < b + \frac{t_{\alpha/2}s}{\sqrt{S_{xx}}} \quad (2.32)$$

Similarly for $A$ we can find a $t$ variate

$$T = \frac{A - \alpha}{s\sqrt{\frac{x^2}{S_{xx}}} \quad (2.33)$$

Hence, the confidence interval $(1 - \alpha) \cdot 100\%$ is

$$a - \frac{t_{\alpha/2}}{\sqrt{\frac{x^2}{S_{xx}}}} < \alpha < a + \frac{t_{\alpha/2}}{\sqrt{S_{xx}}} \quad (2.34)$$

### 2.1.3 Example of Simple Regression

Next we show a simple example of a regression based upon a cooked up model: $\alpha = 5.0, \beta = 2.0, \sigma^2 = 1$. Three realizations are shown in the figure. The statistics of the fits were found using the Regress[] function in MATHEMATICA. It should be possible to guess the meaning of the entries in the table. The data show that in each case there is a significant relationship.
Figure 2.1: Three realizations of the process, $Y = \alpha + \beta x + E$, with $\alpha = 5, \beta = 2, \sigma^2_E = 1$. The solid line in each case is the estimated regression line, and the dashed line (same in all cases) is the true regression line for the process.
Table 2.1: Parameter table for the 3 fits shown in the figure.

<table>
<thead>
<tr>
<th>case</th>
<th>coeff</th>
<th>Est</th>
<th>SE</th>
<th>TStat</th>
<th>PValue</th>
<th>RSS</th>
<th>SSE/3</th>
<th>FStat</th>
<th>PValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>4.649</td>
<td>0.959</td>
<td>4.850</td>
<td>0.0017</td>
<td>68.5</td>
<td>0.835</td>
<td>82.0</td>
<td>0.0028</td>
</tr>
<tr>
<td>1</td>
<td>B</td>
<td>2.617</td>
<td>0.289</td>
<td>9.055</td>
<td>0.0028</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>4.280</td>
<td>2.115</td>
<td>2.024</td>
<td>0.1361</td>
<td>31.78</td>
<td>4.06</td>
<td>7.81</td>
<td>0.068</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>1.783</td>
<td>0.638</td>
<td>2.796</td>
<td>0.0681</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>5.940</td>
<td>2.169</td>
<td>2.738</td>
<td>0.0714</td>
<td>31.71</td>
<td>4.27</td>
<td>7.41</td>
<td>0.0724</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>1.781</td>
<td>0.654</td>
<td>2.723</td>
<td>0.0724</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.2 Apportioning Variance

We now examine the way our errors can be accounted for

\[
SSE = \sum_{i=1}^{n} (y_i - a - bx_i)^2
\]

\[
= \sum (y_i - \bar{y} - b(x_i - \bar{x}))^2
\]

\[
= \sum (y_i - \bar{y})^2 - 2b \sum (x_i - \bar{x})(y_i - \bar{y}) + b^2 \sum (x_i - \bar{x})^2
\]

\[
= S_{yy} - 2bS_{xy} + b^2S_{xx}
\]

\[
= S_{yy} - bS_{xy}
\]

where in the last step we used \( b = S_{xy}/S_{xx} \). Note that if \( b = 0 \)

\[
SSE = S_{yy}
\]

and the errors account for all the variance in \( Y \). We can call

\[
S_{yy} = SST
\]

The Total Sum of Squares of deviations of \( Y \) from its average \( \bar{Y} \). Then

\[
SST = SSE + bS_{xy},
\]

or

\[
SST = SSE + S_{xx}b^2
\]

The second term is conveniently thought of as the Regression Sum of Squares: SSR. Hence

\[
SST = SSE + SSR
\]

The second term is said to ‘explain’ the portion of the variance due to the relationship between \( Y_i \) and \( x_i \). Of course, it may be that \( SSR > 0 \) just
because of chance (sampling) and \( b \) really should be zero (\( \beta = 0 \)). Since the quantity \( B \sqrt{S_{xx}} \) has variance \( \sigma^2 \) and \( B \sqrt{S_{xx}} \) is normally distributed, we have that \( SSR/\sigma^2 \) is \( \chi^2_1 \). We also know that \( SSE/\sigma^2 \sim \chi^2_{n-2} \) and it makes sense that \( SST/\sigma^2 \) would be \( \chi^2_{n-2} \). Under the null hypothesis \( H_0 \) that \( \beta = 0 \), we can assert that

\[
f = \frac{SSR/1}{SSE/(n-2)} = \frac{SSR}{s^2}
\]

is an F statistic: \( F(1, n - 2) \). Hence, we would reject \( H_0 \) at the \( \alpha \) level of significance when \( f > f_\alpha(1, n - 2) \). This is an example of analysis of variance. Note in the ratio the \( \sigma^2 \) was common to the numerator and denominator and therefore cancelled out.

### 2.3 Correlation

It is very important to keep in mind that in all of the above we have used a model, (2.13). Naturally, not all two dimensional situations would follow this model. Frequently, we have to deal with the situation where \( X \) and \( Y \) are both rv’s. In this situation it is far better to deal with a more symmetrical formalism rather than a model like (2.13). [Picture of a scatter plot]. It is easy to imagine physical situations like this one. We define the **correlation coefficient**

\[
\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{xy}}{\sigma_x\sigma_y}
\]

A sample estimate of \( \rho \) is

\[
r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2\sum(y - \bar{y})^2}}
\]

Note that \( r \) and \( \rho \) are symmetric in \( x \) and \( y \). When \( X \) and \( Y \) are independent \( \rho = 0 \) and when \( X = Y \), \( \rho = 1 \); \( X = -Y \) leads to \( \rho = -1 \). Pdfs for these situations are for \( X, Y \) normally distributed, \( \rho = 0 \):

\[
f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\left(\frac{x^2}{2\sigma^2_x} + \frac{y^2}{2\sigma^2_y}\right)}
\]

More generally,

\[
f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \left[ \frac{x^2}{\sigma^2_x} - 2\frac{\rho}{\sigma_x\sigma_y} xy + \frac{y^2}{\sigma^2_y}\right]\right)
\]
Figure 2.2: Equal correlation contours for a process with correlation coefficient $\rho = 0.5$. In this case $\sigma_x = 1, \sigma_y = 2$. 
Figure 2.3: Equal correlation contours for a process with correlation coefficient $\rho = 0.8$. In this case $\sigma_x = 1, \sigma_y = 2$. 
Figure 2.4: Equal correlation contours for a process with correlation coefficient $\rho = 0.0$. In this case $\sigma_x = 1, \sigma_y = 2$. 
Figure 2.5: Scatter diagram for ρ = 0 based upon a sample of 50 points.
Figure 2.6: Scatter diagram for $\rho = 0$ based upon a sample of 500 points.
2.4. SIGNIFICANCE TEST FOR R

In the case of regression we are looking at the conditional pdf \( f(y|x) \):

\[
f(y|x) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left( \frac{-(y - \alpha - \beta x)^2}{2\sigma^2} \right)
\]  (2.47)

Now using the fact that

\[
\beta = \frac{\sigma_x}{\sigma_y} \quad (2.48)
\]

\[
\alpha = \mu_y - \beta \mu_x \quad (2.49)
\]

and

\[
f(x) = \frac{1}{\sqrt{2\pi\sigma_x}} \exp \left( \frac{-(x - \mu_x)^2}{2\sigma_x^2} \right)
\]  (2.50)

and the rule

\[
f(x, y) = f(y|x) f(x)
\]  (2.51)

reconciles the two points of view.

When both \( X \) and \( Y \) are random, we can write

\[
Y = \beta X + E
\]  (2.52)

with \( \beta = \rho \sigma_x / \sigma_y \) (\( \rho \sigma_x / \sigma_y \)) \( |\rho| \leq 1 \). Another point of view is

\[
X = \beta' Y + E
\]  (2.53)

where \( \beta' = \rho \sigma_y / \sigma_x \) (\( \sigma_x / \sigma_y \)). Formally, we can compute the correlation coefficient for the model (2.13).

\[
r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = b \sqrt{\frac{S_{xx}}{S_{yy}}}
\]  (2.54)

We can see that \( r = 0 \) if \( b = 0 \). Also we can show

\[
r^2 = 1 - \frac{SSE}{S_{yy}} = 1 - \frac{SSE}{SST}
\]  (2.55)

Since \( SST > SSE \) we have that \( r^2 < 1 \).

2.4 Significance Test for \( r \)

Remember that \( r \) is only a sample estimate of the true correlation \( \rho \). When we make an estimate of \( \rho \) there will always be sampling errors and these will tend to make us think there is correlation even when there isn’t. Suppose
for example there is no correlation between the normally distributed rv’s
X and Y. Our null hypothesis then is that ρ = 0

\[ H_0 : \rho = 0 \]  \hspace{1cm} (2.56)

Sampling errors can lead to \( r^2 > 0 \). We want to see how \( r \) is distributed
under \( H_0 \). From above we have

\[ r^2 = B^2 \frac{S_{xx}}{S_{yy}} \]  \hspace{1cm} (2.57)

(thinking of the estimator \( B \) as a random variable). Then

\[ 1 - r^2 = \frac{SSE}{S_{yy}} \]  \hspace{1cm} (2.58)

then

\[ \frac{r}{\sqrt{1 - r^2}} = \frac{B \sqrt{S_{xx}}}{\sqrt{SSE}} \]  \hspace{1cm} (2.59)

As we have seen for \( \beta = 0 \) (no correlation) we get

\[ t_{[n-2]} = \frac{B \sqrt{S_{xx}}}{\sqrt{SSE}} \sqrt{n - 2} \]  \hspace{1cm} (2.60)

hence we find that the quantity

\[ \frac{r}{\sqrt{1 - r^2}} \sqrt{n - 2} = \frac{B \sqrt{S_{xx}}}{\sqrt{SSE}} = t_{[n-2]} \]  \hspace{1cm} (2.61)

We can then see if \( r \) is significantly different from zero by checking to see if
the magnitude of the LHS of the last exceeds \( t_{[n-2],\alpha} \) where \( \alpha \) denotes the
level of significance. Note that we lose 2 dof because \( \bar{x} \) and \( \bar{y} \) are determined
from the data. It is remarkable how often a seemingly large value of \( r \) fails
the test because \( n \) is too small!

### 2.4.1 When \( \rho \neq 0 \)

When \( \rho \neq 0 \) in the null hypothesis we can still find out about the expected
distributin of \( r \) but the formulas are more complicated. Under the same
assumptions bout X and Y we find that the variate

\[ z = \frac{1}{2} \ln \left( \frac{1 + r}{1 - r} \right) \]  \hspace{1cm} (2.62)
2.4. SIGNIFICANCE TEST FOR $R$

is approximately normally distributed (Fisher 1921; See Bulmer p225) with mean

$$\mu_z = \frac{1}{2} \ln \left( \frac{1 + \rho}{1 - \rho} \right)$$

(2.63)

and variance $1/(n-3)$. This allows us to conduct tests and find confidence intervals for $\rho \neq 0$.

2.4.2 Correlation between two curves

In the literature one occasionally sees two curves correlated. Let's look critically at this problem [picture of two curves $y_1(t)$ and $y_2(t)$]. Perhaps one is observational data, the other some theory with adjustable parameters in it. We can formally construct a correlation

$$r = \frac{\sum_i (y_1 - \bar{y}_1)(y_2 - \bar{y}_2)i}{\sqrt{\sum_i (y_1 - \bar{y}_1)^2 \sum_i (y_2 - \bar{y}_2)^2}}$$

(2.64)

$$= \frac{S_{12}}{\sqrt{S_{11}S_{22}}}$$

(2.65)

We could now presumably go back and test the null hypothesis that $y_1(t)$ and $y_2(t)$ are uncorrelated. The problem of course is in estimating the number of degrees of freedom. Obviously we have to correct for the fact that $y_2(t)$ the theoretical curve has a few adjustable parameters and subtract one dof for each. There is another problem, however! $y_1(t_1)$ and $y_1(t_2)$ may be correlated and hence we do not have $n$ independent drawings of the variable $y_1(t)$. We shall see later in time series analysis how to find the effective number of independent samples over the range $t_1$ to $t_n$. Only then can we judge the statistical significance of $r$ under $H_0$. Another approach would be to look at

$$\chi^2 = \sum_i \frac{(e_i - \alpha_i)^2}{e_i}$$

(2.66)

the goodness of fit test, but the fact that the $i$ and $i+1$ values are correlated brings us back to the same dof problem.

2.4.3 Max Likelihood and Linear Regression

In computing the coefficients $a$ and $b$ in the linear regression equation $y = a + bx$, we are estimating $\alpha$ and $\beta$ in the system

$$Y = \alpha + \beta x + E$$

(2.67)
where the pdf for the rv $E$ is known ($N(0, \sigma^2)$). The likelihood function for a set of measurements $Y_i$ from the data points $x_i$ is then

$$L(\alpha, \beta) = \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^n e^{-E^2/\sigma^2} \quad (2.68)$$

We have

$$\ln L(\alpha, \beta) = n \ln(\sqrt{2\pi\sigma}) - \sum (y_i - \alpha - \beta x_i)^2 / 2\sigma^2 \quad (2.69)$$

Note now that application of the maximum likelihood criterion, $\partial L/\partial \alpha, \beta = 0$ leads to the same normal equations as the least squares methods. We then have m.l.e. $\hat{A}$ and $\hat{B}$ the same as before. It’s all very comforting.

### 2.5 Multiple Linear Regression

Now we take up the important problem of fitting multivariate data. In this case there are several independent data variables and we wish to find the best plane passing through the data in the sense of least deviations squared. For example

$$y = ax + bz + c \quad (2.70)$$

and we are given triples of data $(Y_i, x_i, z_i), i = 1, \ldots, n$. A model for such a situation is in analogy to (2.13)

$$Y = \alpha x + \beta z + \nu + E \quad (2.71)$$

We shall attempt to find a regression plane to do the job by minimizing the residuals

$$e^2 = \sum_{i=1}^{n} (y_i - ax_i - bz_i - c)^2 \quad (2.72)$$

by taking derivatives as usual

$$\frac{\partial e^2}{\partial a} = \frac{\partial e^2}{\partial b} = \frac{\partial e^2}{\partial c} = 0 \quad (2.73)$$

These three linear equations (Normal equations) can then be solved for $a, b, c$ in exact analogy with the 2 dimensional case already discussed.

To treat the general case we need some fancy notation. Imagine $k$ independent variables $x_\alpha, \alpha = 1, \ldots, k$. There are $n$ realizations of $k$-tuples of observations

$$x_\alpha(i), \ i = 1, \ldots, n. \quad (2.74)$$
2.5. **MULTIPLE LINEAR REGRESSION**

A *response* $y(i)$ is measured for each realization $i$. The regression plane is

$$y(i) = \sum_{\alpha=1}^{k} x_{\alpha}(i)b_{\alpha} + b_{0} \quad (2.75)$$

If one of the variables $x_{\alpha}$ is taken to be unity, then we can absorb the constant term into the sum leading to

$$y(i) = \sum_{\alpha=1}^{k} b_{\alpha}x_{\alpha}(i) \quad (2.76)$$

We are still employing the model

$$Y = \sum_{\alpha} \beta_{\alpha}x^{'\alpha}(i) + E \quad (2.77)$$

More notation:

$$S_{\alpha\alpha} = \bar{x}_{\alpha} \cdot \bar{x}_{\alpha} = \sum_{i=1}^{n} x_{\alpha}(i)x_{\alpha}(i) \quad (2.78)$$

$$S_{y\alpha} = \bar{x}_{\alpha} \cdot \bar{y} = \sum_{i=1}^{n} x_{\alpha}(i)y(i) \quad (2.79)$$

Similarly $S_{\alpha\beta}$, etc; note $S_{\alpha\beta} = S_{\beta\alpha}$.

In analogy to [earlier] the residual sum of squares of errors $SSE$

$$e^{2} = \sum_{i=1}^{n} (y(i) - \sum_{\alpha} b_{\alpha}x_{\alpha}(i))^{2} \quad (2.80)$$

Then

$$\frac{\partial e^{2}}{\partial b_{\beta}} = -2 \sum_{i=1}^{n} n x_{\beta}(y(i) - \sum_{\alpha} b_{\alpha}x_{\alpha}(i)) \rightarrow 0 \quad (2.81)$$

or

$$S_{y\beta} - \sum_{\alpha} b_{\alpha}S_{\alpha\beta} \quad (2.82)$$

The normal equations are then

$$\sum_{\alpha=1}^{k} S_{\beta\alpha}b_{\alpha} = S_{y\beta} \quad (2.83)$$

These are $k$ equations in the $k$ unknowns $b_{\beta}$. (Note $b_{0}$ has already been eliminated through $\bar{y} = b_{0} + \sum b_{\alpha}\bar{x}_{\alpha}$, which can be used after $b_{\alpha}, \alpha > 0$ are
found). The matrix $S_{\alpha\beta}$ can be inverted if it is nonsingular (what does it mean if it is singular?). Hence, we find (if we have a computer with stats software):

$$b_\alpha = \sum_\beta C_{\alpha\beta} S_{\beta\beta}$$  \hspace{1cm} (2.84)

where

$$C_{\lambda\mu} = (S_{\alpha\beta}^{-1})_{\lambda\mu}$$  \hspace{1cm} (2.85)

We shall see that $C_{\lambda\mu}$ is an important matrix.

Now as usual we are faced with the problem of sampling errors. We can look upon $B_\alpha$ as an estimator of $\beta_\alpha$ the true regression coefficient. Hence $B_\alpha$ is an rv.

$$B_\alpha = \sum_\beta C_{\alpha\beta} S_{\beta\beta}$$  \hspace{1cm} (2.86)

Note that $C_{\alpha\beta}$ depends only upon the $x_\alpha$ and is not an rv whereas $S_{\beta\beta}$ is since it is linear in $Y'$. Clearly, then the $B_\alpha$ are normally distributed for large $n$ since they are linear combo's of the $Y(i)$.

$$B_\alpha = \sum_\beta C_{\alpha\beta} \sum_i (x(i) - \bar{x})(Y(i) - \bar{Y})$$  \hspace{1cm} (2.87)

Here $\bar{Y}$ is also an rv. We may relate $Y_i - \bar{Y}$ to $\sum \beta_a x_\alpha + E_i - \bar{E}_i$ and note that Cov($x_\alpha(j)E_i$) = 0.

$$E(B_\alpha) = \sum_{\alpha\beta} C_{\alpha\beta}(x_\alpha(i) - \bar{x})\beta_\gamma(x_\gamma(i) - \bar{x})$$  \hspace{1cm} (2.88)

$$= \sum_{\alpha\beta} C_{\alpha\beta}\beta_\gamma S_{\alpha\gamma}$$  \hspace{1cm} (2.89)

$$= \beta_\alpha \text{ since } C = S^{-1}$$  \hspace{1cm} (2.90)

Hence $B_\alpha$ is an unbiased estimator. Now we wish to examine $V(B_\alpha)$ or more generally Cov($B_\alpha B_\mu$).

$$E\{(B_\alpha - \beta_\alpha)(B_\mu - \beta_\mu)\} = E(B_\alpha B_\mu) - \beta_\mu \beta_\alpha$$  \hspace{1cm} (2.91)

After some work:

$$= C_{\alpha\mu}$$  \hspace{1cm} (2.92)
2.6 Multivariate Normal Distribution

We have \( k \) random variables \( Z_\alpha, \alpha = 1, \ldots, k \), with covariance

\[
\text{Cov}(Z_\alpha Z_\beta) = E((Z_\alpha - \mu_\alpha)(Z_\beta - \mu_\beta)) = \Sigma_{\alpha\beta}
\]

(2.93)

(2.94)

We will make use of the inverse of the covariance matrix: \( \Sigma^{-1} \) and the determinant \( |\Sigma| \). The pdf of a multivariate normal process is

\[
f(z_1, z_2, \ldots, z_k) = \frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}}} \exp \left[ -\frac{1}{2} \sum_{\alpha\beta} (z - \mu_\alpha)(\Sigma^{-1})_{\alpha\beta}(z - \mu_\beta) \right]
\]

(2.95)

We say \( Z_\alpha \sim N(\bar{\mu}, \Sigma) \).

2.6.1 Confidence regions

The estimators \( B_\alpha \) are a multivariate normally distributed vector

\[
B_\alpha \sim N(\bar{\beta}, \sigma^2 C)
\]

(2.96)

and \( C = S^{-1} \)

(2.97)

2.6.2 2 Dimensional Example

\[
Y = \beta_1 x_1 + \beta_2 x_2 + E
\]

(2.98)

or

\[
\hat{y} = b_1 x_1 + b_2 x_2 \text{ regression plane}
\]

(2.99)

1. We have data \((y, x_1, x_2)_1, \ldots, (y, x_1, x_2)_n\)

2. Compute \( S_{11}, S_{12}, S_{22} \); invert to get \( C_{11}, C_{12}, C_{22} \).

3. Estimate \( \hat{\sigma}^2 = SSE/(n - 2 - 1), S_{1y}, S_{2y} \)

4. \( b_1 = \sum_\alpha C_{1\alpha} S_{y\alpha}, b_2 = \sum_\alpha C_{2\alpha} S_{y\alpha} \)

2.6.3 Confidence regions

\[
f(B_1, \ldots, B_k) \sim N(\bar{\beta}, S_{\alpha\beta})
\]

(2.100)

The quadratic form

\[
\sum_{\alpha,\beta} (B_\alpha - \beta_\alpha) S_{\alpha\beta}(B_\beta - \beta_\beta)
\]

(2.101)
is a $\chi^2$ variate with $k$ dof. We can think of the pdf of the errors

$$f(e) = \left(\frac{1}{2\pi}\right)^{k/2} \frac{1}{|C|^{1/2}} \exp(-e^2/2)$$  \hspace{1cm} (2.102)

When $e^2$ is too large we will tend to reject our fit as lousy. Hence, we could think of a volume defined by $e^2 < e_{\text{crit}}^2$ which is the acceptable region [confidence region]. We can say that for $\chi^2 < e_{\text{crit}}^2$, all is well.

2.6.4 2 Dimensional case

$S$ and $C$ (also $b_1$ and $b_2$) have been calculated from the data. Now

$$e^2 = (b_1 - \beta_1)S_{11}(b_1 - \beta_1) + (b_1 - \beta_1)S_{12}(b_2 - \beta_2) + (b_2 - \beta_2)S_{21}(b_1 - \beta_1) + (b_2 - \beta_2)S_{22}(b_2 - \beta_2)$$  \hspace{1cm} (2.103)

$$e^2 = U\beta_1^2 + V\beta_1 + W\beta_1\beta_2 + S\beta_2^2 + R\beta_2 + T$$  \hspace{1cm} (2.104)

Note that $b_1, b_2, S_{ij}$ are computed, $\beta_1, \beta_2$ are free. We could write

where the capital letters stand for constants. For $e^2 = e_{\text{crit}}^2$ we have an ellipse [picture of an ellipse in the $\beta_1, \beta_2$ plane]. Inside the ellipse is the confidence region. If the origin is in the ellipse we must reject our model completely!

2.7 Prediction

Multiple regression is used as a method of prediction in atmospheric science. We generally have a quantity $y$ which is to be predicted from a set of data $(x_1, \ldots, x_k) \rightarrow x_\alpha$. Usually the $x_\alpha$ are in a model sense random variates and so is $Y$. There are also errors $E$. In a linear statistical model we have

$$Y = \sum \beta_\alpha X_\alpha + E$$  \hspace{1cm} (2.106)

a regression plane

$$y = \sum \beta_\alpha x_\alpha$$  \hspace{1cm} (2.107)

is constructed from some realizations of the data by computing the estimates $b_\alpha$ from the formulas. We can then compute confidence regions in the $k$-dimensional space for these coefficients. We can get a feeling for how
Figure 2.7: Three examples of confidence regions in the $b_1, b_2$ plane. Case (a) is one in which coefficient $b_1$ is not significant. Case (b) is one in which neither $b_1$ nor $b_2$ is significant and the model must be entirely rejected. Case (c) is one in which the null hypothesis must be rejected; i.e., the coefficients of the regressions are each significantly different from zero.
good the forecast is by seeing how much of the SST can be reduced by the regression fitting

\[ SST = SSE + SSR \quad (2.108) \]

\[ = S_{yy} \quad (2.109) \]

The fraction of SST explained by the regression

\[ \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = R^2 \quad (2.110) \]

is usually called the multiple regression coefficient (squared). For a discussion of various partial correlation coefficients, see any good book on regression.

In our game \( R^2 \) is usually referred to as the skill, S, and it is extremely important. In climate we frequently want to predict some quantity say the atmospheric pressure at a point and we use as predictors the sea surface temperature months before. All of the changes are small and we only have a small collection of measurements (realizations) to estimate the \( \beta_\alpha \).

To calculate \( p \) at Honolulu, at how many stations in the ocean should we measure the sea surface temperature? (This is an open question). A series of papers has recently appeared worrying about the methodology, most notably


### 2.7.1 More about skill

We are faced with the regression problem (prediction)

\[ \hat{y} = \sum_{\alpha} b_\alpha x_\alpha \quad (2.111) \]

where \( \hat{y} \) is the predicted value of \( y \) (predictand) and the \( x_\alpha \) are the explanatory variables or predictors. Usually \( \bar{x}_\alpha = 0 \) since we work with anomalies (deviations from long term averages). How many cities should we include if we have \( n \) realizations? (\( n \) \( k + 1 \) - tuples of data). clearly if \( n = k \), the \( b_\alpha \) are uniquely determined. When \( n > k \) we try to find \( b_\alpha \) which minimize the errors with the data set in the least squares sense. We can then measure our skill \( S \) by the multiple correlation coefficient \( R \).

\[ R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{(n - k - 1)\hat{\sigma}^2}{S_{yy}} \quad (2.112) \]
We have only \( n \) realizations so our \( R^2(>0) \) might be positive only due to sampling errors. That is, the origin might lie inside the hyperellipsoidal confidence region. The true \( \sigma^2 \) is estimated by

\[
\hat{\sigma}^2 = \frac{SSE}{n - k - 1} \tag{2.113}
\]

where the denominator is clearly the number of dof.

How do we limit the number of explanatory variables when \( n \) is not \( \gg k \)? One method is called ‘screening’ in which \( \hat{\sigma}^2 \) is computed including and excluding variable \( x_j \). If \( \hat{\sigma}^2 \) gets smaller, keep \( x_j \). If \( \hat{\sigma}^2 \) gets larger when \( x_j \) is added, the denominator’s decrease (loss of one dof) is more important than the associated \( SSE \) in the fit. There are various procedures available for reducing the number of variables such as testing the significance of \( b_j \) under the null hypothesis that it is zero. There are various forms of \( F \) test that can be used here. Both Hasselmann and Davis (and North) object to such screening procedures because they are always performed \( a \ posteriori \). That is, you fish through the data for the variables that matter most and if there are a large number, of course, chance will be there to help you (Type I errors). The only safe way to avoid the problem is to start with a \emph{physical} model. This is an \emph{a priori} selection of predictors. Still there is some degree of ambiguity in most practical situations. However, one should develop an \emph{a priori} kind of procedure that does not depend upon any previous examination of the data, correlations, etc. Hasselmann, Davis and others suggest a hierarchy of models based on \emph{Empirical Orthogonal Functions}. 
Exercises
Chapter 3

Empirical Orthogonal Functions

3.1 Matrix Notes

A matrix is a two dimensional array of objects (usually numbers). It takes two indices to locate one of the entries: $M_{kk'}; k, k' = 1, \ldots, n$, where $M$ is an $n \times n$ matrix. In this chapter we will deal mainly such square matrices. The transpose of a matrix, denoted $M^T$ is defined by

$$\{M^T\}_{kk'} = M_{k'k} \quad (3.1)$$

A symmetric matrix is one whose transpose is itself. Multiplication of matrices is accomplished by summing over the adjacent indices:

$$C = AB \quad (3.2)$$

means

$$C_{kk'} = \sum_{k''=1}^{n} A_{kk''} B_{k''k'} \quad (3.3)$$

It is easy to show that

$$(AB)^T = B^T A^T \quad (3.4)$$

A vector $v$ is simply a column matrix requiring a single index: $v_k$. The transpose of $v^T$ or a row matrix. The inner product of two vectors denoted $(u, v)$ is simply the scalar (no indices):

$$(u, v) = u^T v = \sum_{k=1}^{n} u_k v_k \quad (3.5)$$
The length or norm of a vector is \( \sqrt{\langle v, v \rangle} \). A unit vector has its inner product with itself equal to unity. Two vectors are orthogonal if \( \langle u, v \rangle = 1 \).

The outer or the matrix product of two vectors is a matrix
\[
\mathcal{M} = uv^T
\]  
(3.6)

or
\[
\{\mathcal{M}\}_{kk'} = \{u\}_k \{v^T\}_{k'}
\]  
(3.7)

The unit matrix is
\[
\{1\}_{kk'} = \delta_{kk'}
\]  
(3.8)

The inverse of a matrix \( \mathcal{M}^{-1} \) is the matrix which satisfies:
\[
\mathcal{M}^{-1} \mathcal{M} = \mathcal{M} \mathcal{M}^{-1} = 1
\]  
(3.9)

An orthogonal matrix \( \mathcal{O} \) is one whose inverse is equal to its transpose. For an orthogonal matrix then:
\[
\mathcal{O}^T \mathcal{O} = 1
\]  
(3.10)

An orthogonal transformation on a vector is the matrix multiplication of the vector by an orthogonal matrix.
\[
v' = \mathcal{O}v
\]  
(3.11)

It is easy to show that \( \langle v', v' \rangle = \langle v, v \rangle \); i.e., the length is preserved by an orthogonal transformation. If two vectors have a certain inner product \( \langle u, v \rangle \), then its valued is preserved under an orthogonal transformation \( \langle u', v' \rangle = \langle u, v \rangle \). An orthogonal transformation of a matrix \( \mathcal{M} \) is accomplished by
\[
\mathcal{M}' = \mathcal{O} \mathcal{M} \mathcal{O}^T
\]  
(3.12)

The orthogonal transformation of a product of two matrices is easily shown to be the product of the transforms:
\[
(AB)' = \mathcal{O}ABO^T = \mathcal{O}AO^T \mathcal{O}BO^T = A'B'
\]  
(3.13)

A set of \( n \) vectors \( e^{(s)}; s = 1, \ldots n \), are said to span a space of dimension \( n \) if any vector in the space can be written as a linear combination of the set. A particularly useful set are those which are orthonormal, i.e.,
\[
e^{(s)}e^{(s')} = \delta_{ss'}
\]  
(3.14)

They are orthogonal and of unit length.
3.1. **MATRIX NOTES**

3.1.1 **Eigenvectors**

Consider an $n \times n$ symmetric matrix $\mathcal{M}$. Is it possible that for some vectors in the space a multiplication by $\mathcal{M}$ on a vector $\mathbf{u}$ yields a vector that is parallel to (=a scalar multiple of) $\mathbf{u}$?

$$\mathcal{M}\mathbf{u} = \lambda \mathbf{u} \quad (3.15)$$

This can only be so if

$$|\mathcal{M} - \lambda \mathbf{1}| = 0 \quad (3.16)$$

where the $|\cdot|$ indicates the determinant. The last expression is an $n$ degree polynomial set equal to zero. It will have $n$ roots, corresponding to $n$ solutions to the system of equations (3.15). This means there are $n$ vectors satisfying (3.15). Special care must be devoted to the eigenvectors corresponding to the cases in which two or more eigenvalues are equal. For the moment suppose the eigenvalues are strictly distinct from one another. Then we can proceed with an important proof. Consider (3.15) for a specific eigenvalue $\lambda_s$ and its eigenvector $\mathbf{e}^{(s)}$. Then multiply from the left with the transpose of another of the eigenvectors

$$(\mathbf{e}^{(s')}, \mathcal{M}\mathbf{e}^{(s)}) = \lambda_s (\mathbf{e}^{(s')}, \mathbf{e}^{(s)}) \quad (3.17)$$

We can write the same equation with $s$ and $s'$ interchanged.

$$(\mathbf{e}^{(s')}, \mathcal{M}\mathbf{e}^{(s)}) = \lambda_{s'} (\mathbf{e}^{(s')}, \mathbf{e}^{(s)}) \quad (3.18)$$

Next note that

$$(\mathbf{e}^{(s')}, \mathcal{M}\mathbf{e}^{(s)}) = (\mathcal{M}\mathbf{e}^{(s')}, \mathbf{e}^{(s)}) \quad (3.19)$$

where we have used the symmetry of $\mathcal{M}$. Subtracting the two eigen equations and making use of the last relationship, we find

$$(\lambda_s - \lambda_{s'}) (\mathbf{e}^{(s')}, \mathbf{e}^{(s)}) = 0 \quad (3.20)$$

which means that if the eigenvalues are distinct, the eigenvectors are mutually orthogonal. Since there are $n$ of them, they span the $n$ dimensional space. It is convenient to normalize the eigenvectors to unit length. Any vector in the space can be expanded into a linear combination of the eigenvectors

$$\mathbf{v} = \sum_{s=1}^{n} V_s \mathbf{e}^{(s)} \quad (3.21)$$

or

$$\{\mathbf{v}\}_k = v_k = \sum_{s=1}^{n} V_s \{\mathbf{e}^{(s)}\}_k \quad (3.22)$$
CHAPTER 3. EMPIRICAL ORTHOGONAL FUNCTIONS

with the coefficients, $V_s$ given by

$$V_s = \sum_{k=1}^{n} e_k^{(s)} \{v \}_{k} = (e^{(s)}, v)$$ (3.23)

Consider the matrix whose rows are the eigenvectors

$$\mathcal{O}_{s,k} = e_k^{(s)}$$ (3.24)

Clearly $\mathcal{O}$ is an orthogonal matrix since $\mathcal{O}\mathcal{O}^T = 1$. Now consider the symmetric matrix $\mathcal{M}$ whose eigenvectors are the $e^{(s)}$. The orthogonal transformation $\mathcal{O}$ on $\mathcal{M}$ yields

$$(\mathcal{O}\mathcal{M}\mathcal{O}^T)_{ss'} = (\mathcal{M}')_{ss'} = \lambda_s \delta_{ss'}$$ (3.25)

In other words the special orthogonal transformation (3.24) is the transformation to a new coordinate system in which $\mathcal{M}'$ is diagonal with the eigenvalues $\lambda_s$ along its diagonal.

3.2 Definition of EOFs

Suppose the explanatory variables $X_\alpha, \alpha = 1, \ldots, k$ are random variables but possibly correlated with each other. We might imagine as a mnemonic that our data come from stations in Greece, hence are labeled by the Greek letters. As a model take them to be multivariate normal $N(\mathbf{\beta}, \Sigma)$. Note that if the covariance matrix $\Sigma$ is diagonal (no cross-correlations) then the argument of the exponential in (2.95) simplifies to a sum of squares. (note if $\mathbf{S} = \Sigma$ is diagonal, so is $\mathbf{C} = \mathbf{S}^{-1}$). EOFs are just the linear combinations of the $X_\alpha$ that diagonalize the covariance matrix. We proceed to find them.

Let

$$\Sigma_{\alpha\beta} = \text{Cov}(X_\alpha, X_\beta)$$ (3.26)

We often deal with the sample covariance matrix:

$$\hat{\Sigma}_{\alpha\beta} = \frac{\Sigma_{\alpha\beta}}{n} = \frac{1}{n} \sum_{i=1}^{n}(X_\alpha - \bar{X}_\alpha)(X_\beta - \bar{X}_\beta)$$ (3.27)

which is based upon a finite set of data. But for now imagine that we have enough data to have a very good estimate of $\Sigma$. We will return to the question of sampling errors later.

Since $\Sigma_{\alpha\beta}$ (or $\hat{\Sigma}_{\alpha\beta}$) is symmetric, we can do lots of things with it. For example, it has eigenvectors, $\mathbf{e}^{(s)}, s = 1, \ldots, k$. (Each eigenvector is labeled by the index $s$.)

$$\sum_{\beta} \Sigma_{\alpha\beta} e_\beta^{(s)} = \lambda_s e_\alpha^{(s)}$$ (3.28)
3.2. DEFINITION OF EOFS

It is easy to show that the \( \vec{e}(s) \), \( \vec{e}(s') \) are orthogonal. Suppose \( \lambda_s \neq \lambda_{s'} \):

\[
\sum_{\beta} \Sigma_{\alpha\beta} e_{\beta}(s') = \lambda_{s'} e_{\alpha}(s')
\]  

(3.29)

Multiply (3.28) on the left by \( e_{\alpha}(s') \), sum on \( \alpha \); now multiply (3.29) on the left by \( e_{\alpha}(s) \), sum. Next subtract the two expressions from each other. The LHS will cancel due to the symmetry of \( \Sigma \), leaving

\[
(\lambda_s - \lambda_{s'}) \sum_{\alpha} e_{\alpha}(s) e_{\alpha}(s') = 0
\]  

(3.30)

Since \( \lambda_s \neq \lambda_{s'} \), we have

\[
\vec{e}(s) \cdot \vec{e}(s') = \delta_{ss'}
\]  

(3.31)

(the eigenvectors have been normalized to unit length). [for the moment ignore the case where \( \lambda_s = \lambda_{s'} \) which does occur occasionally]. The predictor variables can be expanded into this convenient basis set of unit vectors.

\[
X_{\alpha} = \sum_{s=1}^{k} A_s e_{\alpha}(s)
\]  

(3.32)

Since the \( e_{\alpha}(s) \) are determined and \( X \) is a rv, it follows that the \( A_s \) are rv’s. We can use the orthogonality of the \( \vec{e}(s) \) to compute the expansion coefficients:

\[
A_s = \sum_{\alpha} X_{\alpha} e_{\alpha}(s) = \bar{X} \cdot \vec{e}(s)
\]  

(3.33)

We repeat that the \( A_s \) may be thought of as rv’s , while the \( \vec{e}(s) \) are fixed, since the latter are properties of the covariance matrix, which is derive from the population statistics. We can easily show that

\[
E(A_s A_{s'}) = \sum_{\alpha} \sum_{\beta} e_{\alpha}(s)e_{\beta}(s') E(X_{\alpha}X_{\beta})
\]  

(3.34)

\[
= \sum_{\alpha} \sum_{\beta} e_{\alpha}(s) \Sigma_{\alpha\beta} e_{\beta}(s')
\]  

(3.35)

\[
= \sum_{\alpha} e_{\alpha}(s) \lambda_{s'} e_{\alpha}(s')
\]  

(3.36)

\[
= \lambda_s \delta_{ss'}
\]  

(3.37)

Where we have used \( E(X_{\gamma}) = 0 \). Hence, the variance of \( A_s \) is just \( \lambda_s \), the corresponding eigenvalue and the rv’s \( A_s \) are uncorrelated! Another twist
is to look at the variance of $x_\alpha$

$$V(X_\alpha) = \sum_s V(A_s)e_\alpha^2(s)$$  \hspace{1cm} (3.38)

$$= \sum_s \lambda_se_\alpha^2(s)$$  \hspace{1cm} (3.39)

Now consider the *average* variance of the $X_\alpha$ which is proportional to the sum:

$$\sum_\alpha V(X_\alpha) = \sum_{s=1}^k \lambda_s$$  \hspace{1cm} (3.40)

The last two equations give us an idea how the variance is apportioned among the various EOFs. It is customary to label the EOFs such that the eigenvalues are in order of size, i.e., $\lambda_1 > \lambda_2 \ldots > \lambda_k$. Hence in the last equation the last terms in the sum contribute least to the total variance. The $\lambda_s$ can be represented as a ‘spectrum’ [bar graph of $\lambda_s$ versus $s$]. By definition the graph is monotonically decreasing. The area (sum of vertical line lengths) out to $s_0$ tells the variance accounted for by the subsum to $s_0$. It is possible to show that no other orthogonal basis set with $s_0$ members can account for more variance in the sum.

### 3.3 Example in 2D

This can be easily seen in a 2 dimensional example. Suppose $X_1$ and $X_2$ are rv’s $\sim N(0,\Sigma)$ with pdf [pic of tilted ellipse in $x_1, x_2$ plane]. The transformation (3.33) carries us to a new set of rv’s $A_1, A_2$. The transformation is orthogonal so that it amounts to a rotation of axes:

$$(X_1, X_2) \xrightarrow{\text{rotate}} (A_1, A_2)$$  \hspace{1cm} (3.41)

$$A_1 = X_1 \cos \theta + X_2 \sin \theta$$  \hspace{1cm} (3.42)

$$A_2 = -X_1 \sin \theta + X_2 \cos \theta$$  \hspace{1cm} (3.43)

Next we find the angle $\theta$. Note from (3.33)

$$e_1(1) = \cos \theta \quad , \quad e_2(1) = \sin \theta$$  \hspace{1cm} (3.44)

$$e_1(2) = -\sin \theta \quad , \quad e_2(2) = \cos \theta$$  \hspace{1cm} (3.45)

We can find

$$X_1 = A_1 \cos \theta - A_2 \sin \theta$$  \hspace{1cm} (3.46)

$$X_2 = A_1 \sin \theta + A_2 \cos \theta$$  \hspace{1cm} (3.47)
The covariance in 2 dimensions is
\[
\Sigma = \begin{pmatrix}
\sigma_1^2 & \sigma_1 \sigma_2 \rho \\
\sigma_1 \sigma_2 \rho & \sigma_2^2
\end{pmatrix}
\]

We get for EOF1:
\[
\begin{align*}
\sigma_1^2 \cos \theta + \sigma_1 \sigma_2 \rho \sin \theta &= \lambda_1 \cos \theta \\
\sigma_1 \sigma_2 \rho \cos \theta + \sigma_2^2 \sin \theta &= \lambda_1 \sin \theta
\end{align*}
\]

(3.48)  (3.49)

divide these two to eliminate \( \lambda_1 \), also divide num and denom by \( \cos \theta \):
\[
\tan \theta = \frac{\sigma_1 \sigma_2 \rho + \sigma_2^2 \tan \theta}{\sigma_1^2 + \sigma_1 \sigma_2 \rho \tan \theta}
\]

(3.50)

use \( \tan \theta \equiv t \):
\[
\sigma_1^2 t + \sigma_1 \sigma_2 \rho t^2 = \sigma_1 \sigma_2 \rho + \sigma_2^2 t
\]

(3.51)

Let
\[
w = \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1 \sigma_2 \rho}
\]

(3.52)

Then
\[
\tan^2 \theta + w \tan \theta - 1 = 0
\]

(3.53)

Since the covariance matrix for \((A_1, A_2)\) is diagonal and the \(A_1, A_2 \sim N(0, \Sigma_A)\)
\[
\begin{align*}
(\Sigma_A)_{ss'} &= \text{Cov}(A_s, A_{s'}) = \lambda_s \delta_{ss'} \\
(\Sigma_A^{-1})_{ss'} &= \frac{1}{\lambda_s} \delta_{ss'}
\end{align*}
\]

(3.54)  (3.55)

The pdf is
\[
f(A_1, A_2) = \frac{1}{2\pi} \left( \frac{1}{\lambda_1 \lambda_2} \right)^{1/2} \exp \left( -\frac{A_1^2}{2\lambda_1} - \frac{A_2^2}{2\lambda_2} \right)
\]

(3.56)

The rotation of axes is clearly to the principal axes of the ellipse.

Now it is obvious that the rotation carries us to a set of rv’s \(A_1, A_2\) which are orthogonal and \(A_1\) carries more variance than any other possible choice of rotated axes. For three dimensions we have an ellipsoid and clearly the same conclusion follows.

We can find the eigenvalues \(\lambda_1, \lambda_2\) by setting the determinant of \(\Sigma - \lambda I\) to zero,
\[
\begin{vmatrix}
\sigma_1^2 - \lambda & \sigma_1 \sigma_2 \rho \\
\sigma_1 \sigma_2 \rho & \sigma_2^2 - \lambda
\end{vmatrix} = 0
\]
which is a quadratic in \( \lambda \) with two roots, one for \( \lambda_1 \) and one for \( \lambda_2 \). The formal solution is

\[
\lambda_{\pm} = \frac{\sigma_1^2 + \sigma_2^2}{2} \pm \frac{1}{2} \sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\sigma_1^2\sigma_2^2\rho} \quad (3.57)
\]

The two roots become degenerate if the radical vanishes:

\[
(\sigma_1^2 - \sigma_2^2)^2 + 4\sigma_1^2\sigma_2^2\rho = 0 \quad (3.58)
\]

This latter can only happen if \( \sigma_1^2 = \sigma_2^2 \) and \( \rho = 0 \) simultaneously. In other words if both original variates \( X_1 \) and \( X_2 \) both have the same variance and they are uncorrelated. In this case the ellipse becomes a circle and the angle of rotation is undetermined. This may seem a trivial case but it frequently happens that two eigenvalues are degenerate. For example, we may be dealing with a three variate problem and the corresponding surface is an ellipsoid of revolution about one of the three axes. In performing the three dimensional rotation to the principal axes, we find that after one rotation to the symmetry axis corresponding to the \( z \) axis, the rotation to a proper set of axes in the \( x - y \) plane is undetermined.

It is possible to show that if two eigenvalues in the list are equal (degenerate) we can find two linearly independent vectors in the subspace, but any linear combination of these will also be an eigenvector. Let \( \vec{e}(1) \) and \( \vec{e}(2) \) be two linearly independent vectors belonging to the same eigenvalue \( \lambda_0 \). Then \( \alpha\vec{e}(1) + \beta\vec{e}(2) \) is also an eigenvector of \( \Sigma_{\gamma\mu} \) corresponding to eigenvalue \( (\alpha + \beta)\lambda_0 \). Hence, we can always pick a linear combination within the subspace which are mutually orthogonal. This is exactly what happens in the 2D case when \( \sigma_1 = \sigma_2 \) and \( \rho = 0 \).

### 3.4 Generating Random Vectors

Suppose we want to generate a sequence of realizations of a random vector \( X_\alpha \) with certain properties, say \( \sim N(0, \Sigma) \). One way is to use the EOFs. By finding the eigenvectors \( e_\alpha(s) \) of \( \Sigma_{\alpha\beta} \). The vectors may be written:

\[
X_\alpha = \sum_s A_s e_\alpha(s) \quad (3.59)
\]

The \( A_s \) are uncorrelated

\[
\langle A_s A_{s'} \rangle = \lambda_s \delta_{ss'} \quad (3.60)
\]

Hence once the \( e_\alpha(s) \), \( \lambda_s \) are known, we can use a random number generator to produce Gaussian random numbers for the \( A_s \) with variance \( \lambda_s \). We do this for each \( s \) over all the EOFs in (3.59) and then construct a realization for \( X_\alpha \).
### 3.5 Sampling Errors for EOFs

In practice we usually have only a sample covariance matrix $\hat{\Sigma}_{\alpha\beta}$. The covariance matrix is symmetric hence all the theory above goes through except that we do not have exactly the correct directions $\vec{e}(s)$ and eigenvalues $\lambda_s$. Instead we have sample estimates of these: $\hat{\vec{e}}(s)$ and $\hat{\lambda}_s$.

### 3.6 Regression with EOFs

In examining the system

$$Y = \sum_{\alpha=1}^{k} \beta_\alpha X_\alpha + E$$ (3.61)

where $\text{Cov}(X_\alpha, E) = 0$, $X_\alpha \sim N(0, \Sigma)$, we may find it useful to go to EOFs.

$$X_\alpha = \sum_{s=1}^{k} a_s e_\alpha(s)$$ (3.62)

$$Y = \sum_{\alpha, s} \beta_\alpha e_\alpha(s) A_s + E$$ (3.63)

$$= \sum_{s} \beta'_s A_s + E$$ (3.64)

The estimators for $\beta'_s, B'_s$ will be $G(\hat{\beta}'_s, \Sigma_{B'})$, where

$$(\Sigma_{B'})_{st} = \frac{\sigma^2}{\lambda_s} \delta_{st}$$ (3.65)

In other words, the $B'_s$ correspond to the principal axes of the ellipse shown earlier.

The main advantage in going to the EOF representation is that it provides some advantage in reducing the number of predictors. We can claim, for example, that if above a level $k*$ in the sum, only measurement errors and sampling errors dominate, then we can truncate the series by an objective criterion. Hence, the procedure might be as follows (Hasselmann, 1979)

1. Find the sample EOFs. (eigenvectors of the $S_{\alpha\beta}$.

2. Proceed to get the $b'/s$ as estimates of $\beta'_s$ above by keeping a certain number of terms in the expansion. i.e., $b_1, \ldots, b_{k*}, k* < k$. 

3. We can now have a hierarchy of models for different choices of \( k^* \). Note that for each choice of \( k^* \) the various \( b_s, s = 1, \ldots, k^* \) will turn out differently, so they must be recomputed each time.

4. For each choice \( k^* \) we compute confidence regions at some preassigned level, say 95\%. As we increase \( k^* \), the confidence region (error bars) will increase until it eventually includes the origin in \( \beta_s \) space. Clearly the optimum \( k^* \) is smaller than this.

3.6.1 Principal predictors (Davis, 1977)

One could take a different approach than the one just mentioned. EOFs tell us an optimum and compact way of representing the \( X_{\alpha}, \alpha = 1, \ldots, k \); they are a basis set \( e_\alpha(s), s = 1, \ldots, k \), which transform us to another set of rv’s \( A_s, s = 1, \ldots, k \) which are uncorrelated and which account for the most variance in \( X = \sum_\alpha X_\alpha \), one by one. The basis set \( e_\alpha(s) \) are also orthogonal and complete. On the other hand, we really want to know about the predictand \( Y \). Usually in meteorology \( Y \) is not just a number but a field, which we represent as a vector having values at the grid points

\[
Y_\lambda = \sum_\alpha b_\alpha^\lambda X_\alpha + E \tag{3.66}
\]

\[
\alpha = 1, \ldots, k \tag{3.67}
\]

\[
\lambda = 1, \ldots, l \tag{3.68}
\]

For example, after going to the EOF representation for the \( X_\alpha \), we may find the EOF_2 contributes more to the variance of \( Y = \sum_\lambda Y_\lambda \) than EOF_1 does. The simplest cure might be to examine the EOF_\alpha = e_\alpha(s) one at a time for their statistical significance and throw out the bad ones. Hasselmann (1979) rejects this procedure as \textit{a posteriori} and produces an example where there is no correlation but by “fishing” we can find some vectors that are apparently statistically significant, but they are in fact type I errors. Hence, the principal predictor idea is probably equivalent to the other screening procedures which are all suspect because of their \textit{a posteriori} difficulty.

3.7 Why EOFs in Atmospheric Science?

Maddala in \textit{Econometrics} devotes only a couple of pages to principal components (=EOFs). He says they are rarely found to be useful. In atmospheric or oceanic sciences there is a good reason why they are useful. Usually the predictor \( X_\alpha \) is a field such as temperature, \( T(r) \). From one realization to
the next (long internal, or random observing time), $T(r)$ may be thought of as a random field. The index $\alpha$ corresponds to grid points $r_\alpha$:

$$X_\alpha = T(r_\alpha)$$ (3.69)

[Note we are usually looking at anomalies, i.e., $E(T(r_\alpha))$ is subtracted out. Hence, $E(T(r_\alpha)) = \langle T(r_\alpha) \rangle = 0$.]

Since the fields tend to be smooth, there will be a strong correlation between the field evaluated at neighboring grid points, $T(r_1), T(r_2)$. In some sense the two are redundant in the regression sense, because they carry almost the same information. The field might be expanded into a set of orthogonal functions $g_n(r)$

$$\int_A g_n(r)g_m(r)\, dr = \delta_{nm}$$ (3.70)

$$T(r) = \sum_{n=0}^{\infty} T_n g_n(r)$$ (3.71)

The rv coefficients are given by

$$T_n = \int_A T(r)g_n(r)\, dr$$ (3.72)

For each realization $T(r)$, we can compute the coefficients $T_n$ through the last equation. Is there any basis set $f_n(r)$ such that the corresponding $T_n$ are uncorrelated? The answer is yes. Consider the generalization of $S_{\alpha \beta}$.

$$K(r, r') = \langle T(r), T(r') \rangle$$ (3.73)

the covariance kernel. We can construct the continuous analog of the matrix eigenvalue equation

$$\int_A K(r, r')f_s(r')\, dr' = \lambda_s f_s(r), \ s = 1, 2, \ldots$$ (3.74)

The $f_s(r)$ are the eigenfunctions of the kernel $K(r, r')$. Note

$$K(r, r') = K(r', r)$$ (3.75)

It is easy to show that if $\lambda_s \neq \lambda_{s'}$,

$$\int_A f_s(r)f_{s'}(r)\, dr = \delta_{ss'}$$ (3.76)
(analogous to the corresponding vector equation). In fact, all of our finite
dimensional results carry over to the continuous field case. If the \( f_n(r) \) are
used as a basis set
\[
T(r) = \sum_{n=0}^{\infty} T_n f_n(r)
\]
(3.77)
where the \( f_n \) are the eigenfunctions of the covariance kernel, the expansion
is called a Karhunen-Loeve expansion.

In practice one cannot have continuous realizations of the field \( T(r) \), but
must be satisfied with station data \( T(r_i) \). Hence the eigenvalue problem
must be replaced by an approximating sum
\[
\sum_i K(r_j, r_i) f_s(r_i) (\Delta A)_{1/2}^{i/2} = \lambda_s f_s(r_j)
\]
(3.78)
Multiply on the left and rearrange
\[
\sum_i (\Delta A)_{1/2}^{i/2} K_{ij} (\Delta A)_{1/2}^{j/2} \cdot f_s(i) = \lambda_s (\Delta A)_{1/2}^{1/2} f_s(j)
\]
(3.79)
Now identify
\[
e_i(s) = (\Delta A)_{1/2}^{1/2} f_s(i)
\]
(3.80)
\[
\Sigma_{ij} = (\Delta A)_{1/2}^{1/2} K(r_i, r_j) (\Delta A)_{1/2}^{1/2}
\]
(3.81)
and we are back to the finite dimensional problem. Note that the “metric”
factors \( (\Delta A)_{1/2}^{1/2} \) cause \( \Sigma_{ij} \) to be symmetric. The factors \( (\Delta A)_{1/2}^{1/2} \) can be
chosen somewhat arbitrarily and, of course, if they are large there will be
errors introduced. [picture with polygons surrounding grid points].

3.8 Analytical Examples of EOFs

Sometimes we can find an analytical solution to the integral form of the
eigenvalue problem. I know of a few examples:

1. Rotational invariance on a circle. In this case \( K(r, r') = K(\theta, \theta') = \langle T(\theta)|T(\theta') \rangle \). Rotational invariance means \( \langle T(\theta) \rangle = 0 \) and
\[
K(\theta, \theta') = K(|\theta - \theta'|)
\]
(3.82)
i.e., the correlation between \( T \) at \( \theta \) and \( T \) at \( \theta' \) depends only upon the
separation \( |\theta - \theta'| \). We have
\[
\int_0^{2\pi} K(|\theta - \theta'|) f_n(\theta') d\theta' = \lambda_n f_n(\theta)
\]
(3.83)
which can be solved by expanding into the Fourier Series (more on these in a later chapter):

\[
K(\theta - \theta') = \frac{K_0}{\sqrt{2\pi}} + \sum_{n=1}^{\infty} K_n \frac{\cos n(\theta - \theta')}{\sqrt{\pi}} \tag{3.84}
\]

Only the cosine terms of the last sum appear since \(K(\cdot)\) is an even function of its argument. We have made use of a few properties of trigonometric integrals:

\[
\int_0^{2\pi} \frac{\sin n\theta \sin m\theta}{\sqrt{\pi}} \sqrt{\pi} \ d\theta = \delta_{nm} \tag{3.85}
\]

\[
\int_0^{2\pi} \frac{\cos n\theta \cos m\theta}{\sqrt{\pi}} \sqrt{\pi} \ d\theta = \delta_{nm} \tag{3.86}
\]

\[
\int_0^{2\pi} \frac{\sin n\theta \cos m\theta}{\sqrt{\pi}} \sqrt{\pi} \ d\theta = 0 \tag{3.87}
\]

where in each of the above \(n\) and \(m\) are positive integers. We also have

\[
\cos(\theta_1 + \theta_2) = \cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2 \tag{3.88}
\]

By using these last properties in the eigenvalue equation we can try the trial eigenfunctions

\[
f_0(\theta) = \frac{1}{\sqrt{2\pi}} \tag{3.89}
\]

\[
f_n^{(+)}(\theta) = \frac{\cos n\theta}{\sqrt{\pi}}, \ n = 1, 2, \ldots \tag{3.90}
\]

\[
f_n^{(-)}(\theta) = \frac{\sin n\theta}{\sqrt{\pi}}, \ n = 1, 2, \ldots \tag{3.91}
\]

Where the superscript \((\pm)\) has been used to distinguish the even from the odd function components. From the identities we easily see that

\[
\int_0^{2\pi} f^{(i)}(\theta)f^{(j)}(\theta) \ d\theta = \delta_{nm} \delta_{ij} \tag{3.92}
\]

By insertion we also find that

\[
\int_0^{2\pi} K(\theta, \theta') f_n^{(\pm)}(\theta') \ d\theta' = K_n f_n^{(\pm)}(\theta) \tag{3.93}
\]

Hence

\[
\lambda_n = K_n, \tag{3.94}
\]
and for \( n > 0 \), both \( f_n^{(+)}(\theta) \) and \( f_n^{(-)}(\theta) \) have the same eigenvalue \( K_n \). We say the eigenvalues are doubly degenerate. For \( n = 0 \) there is only a single eigenfunction, the constant, \( f_0(\theta) = 1/\sqrt{2\pi} \) with eigenvalue \( K_0 \).

Next suppose we wanted to represent a realization of the random function \( T(\theta) \). We can expand

\[
T(\theta) = \sum_{n=0}^{\infty} \sum_{i=\pm} T_n^{(i)} f_n^{(i)}(\theta) \tag{3.95}
\]

where the coefficients can be computed from the orthnormality

\[
T_n^{(i)} = \int_{0}^{2\pi} T(\theta)f_n^{(i)}(\theta) \, d\theta \tag{3.96}
\]

The coefficients \( T_n^{(i)} \) are real random variables. We can take their means to be zero (consider anomalies only). Consider next the covariance between to coefficients

\[
\langle T_n^{(i)} T_m^{(j)} \rangle = \int_{0}^{2\pi} \int_{0}^{2\pi} f_n^{(i)}(\theta)f_m^{(j)}(\theta') \langle T(\theta)T(\theta') \rangle \, d\theta \, d\theta' \tag{3.97}
\]

\[
= \int_{0}^{2\pi} f_n^{(i)}(\theta)K_n f_m^{(j)}(\theta) \, d\theta \tag{3.98}
\]

\[
= K_n \delta_{nm} \delta_{ij} \tag{3.99}
\]

This means that the random coefficients \( T_n^{(i)} \) are uncorrelated except with themselves.

Next look at the variance at a point

\[
\langle T(\theta)^2 \rangle = \sum_{n,i} \sum_{m,j} \langle T_n^{(i)} T_m^{(j)} \rangle f_n^{(i)}(\theta) f_m^{(j)}(\theta) \tag{3.100}
\]

\[
= \sum_n K_n \sum_i (f_n^{(i)}(\theta))^2 \tag{3.101}
\]

\[
= \frac{1}{\pi} \sum_n K_n \tag{3.102}
\]

where in the last step we made use of \( \sin^2 \phi + \cos^2 \phi = 1 \). The point of the last exercise is that the variance at a point is the sum of variances from all the EOFs (no cross terms).

In this section it was demonstrated that in some cases symmetry alone imposes the shapes of the EOFs. The only information content of the random function \( T(\theta) \) is the eigenvalue spectrum \( K_n \). For each \( n > 0 \) there
are two contributing EOFs (doublet) with the same eigenvalue (variance) and for \( n = 0 \) we have a singlet.

To simulate a single realization of such a (gaussian) random function whose spectrum is known, we can use a gaussian random number generator with variance \( K_n \). We simply draw the coefficients \( T_n^{(i)} \) from the random generator according to their eigenvalues (variances) and construct a realization from the expansion.

White spatial noise corresponds to all the eigenvalues \( K_n \) being equal.

### 3.8.1 Rotational invariance on the sphere

The correlation depends only upon the great circle distance or only upon

\[
\mathbf{r} \cdot \mathbf{r}' = \cos \Phi
\]  

(3.103)

where \( \Phi \) is the opening angle between \( \mathbf{r} \) and \( \mathbf{r}' \), unit vectors pointing from the center of the sphere to the two points in question.

\[
K(\mathbf{r}_1, \mathbf{r}_2) = K(\mathbf{r} \cdot \mathbf{r}') = K(\cos \Phi)
\]  

(3.104)

\[
= \sum_{n=0}^{\infty} \frac{(2n + 1)K_n}{4\pi} P_n(\cos \Phi)
\]  

(3.105)

where \( P_n(\cos \Phi) \) are the Legendre Polynomials (the factor \( (2n + 1)/4\pi \) is conventional). [See North and Cahalan, 1981.] A key identity is the Addition theorem for spherical harmonics (Arfken, 1985)

\[
P_n(\mathbf{r} \cdot \mathbf{r}') = \frac{4\pi}{2n + 1} \sum_{m=-n}^{n} Y_n^m(\mathbf{r})Y_n^m(\mathbf{r}')
\]  

(3.106)

where the \( Y_n^m(\mathbf{r}) \) are the complex spherical harmonics defined by

\[
Y_n^m(\mathbf{r}) = (-1)^m \sqrt{\frac{2n + 1}{4\pi} \frac{(n - m)!}{(n + m)!}} P_n^m(\cos \theta)e^{im\phi}
\]  

(3.107)

where the polar and azimuthal angles of the unit vector \( \mathbf{r} \) are \( \theta \) and \( \phi \) respectively. The functions \( P_n^m(\cos \theta) \) are the Associated Legendre functions.

The spherical harmonics have the properties:

\[
\int \int Y_n^m(\mathbf{r})Y_n^{m'}(\mathbf{r})d^2\mathbf{r} = \delta_{nn'}\delta_{mm'}
\]  

(3.108)

\[
\sum_{n=0}^{\infty} \sum_{m=-n}^{n} Y_n^m(\mathbf{r})Y_n^{m*}(\mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}')
\]  

(3.109)
CHAPTER 3.  EMPIRICAL ORTHOGONAL FUNCTIONS

As with the circle we can find the eigenfunctions (EOFs) for this symmetric problem using the identities above. We find that the EOFs are just the spherical harmonics. The $n$th degree EOF is $2n + 1$-fold degenerate with eigenvalue $K_n$. As before white spatial noise corresponds to all the eigenvalues $K_n$ being equal.

Note that in this case we have for the first time encountered complex eigenfunctions $Y_m^m(\hat{r})$. We could have used the real spherical harmonics $Y_m^\pm(\hat{r})$ corresponding to sines and cosines as before.

Similarly we could have used the complex basis on the circle:

$$f_n(\theta) = \frac{e^{in\theta}}{\sqrt{2\pi}}, \quad n = 0, \pm 1, \pm 2, \ldots$$  (3.110)

In this case we must use the orthogonality conditions

$$\int_0^{2\pi} f_n(\theta)f_m(\theta)^* \, d\theta = \delta_{nm}$$  (3.111)

and the completeness relation

$$\sum_{n=-\infty}^{\infty} f_n(\theta)f_n(\theta')^* = \delta(\theta - \theta')$$  (3.112)

3.8.2 Translation invariance along a line

Suppose the function $T(t)$ has statistics which are homogeneous on the $t$ axis, $-\infty < t < \infty$,

$$\langle T(t) \rangle = 0, \text{ all } t$$  (3.113)

and

$$\langle T(t)T(t') \rangle = K(|t - t'|)$$  (3.114)

which states that the correlation between the field’s value at $t$ and $t'$ depends only on the separation (lag) between the two times. Now the Karhunen-Loeve equation (using $g_f$ as the EOF) is:

$$\int_{-\infty}^{+\infty} K(t - t')g_f(t') \, dt' = \gamma_f g_f(t)$$  (3.115)

may be solved by the choice

$$g_f(t) = e^{-2\pi if t}$$  (3.116)

since we may write

$$K(\tau) = \int_{-\infty}^{+\infty} \hat{K}(f')e^{2\pi if'\tau} \, df'$$  (3.117)
3.9. MECHANICAL SYSTEMS

(Fourier Integral Representation); then the LHS of the eigenvalue equation becomes

\[
\int_{-\infty}^{+\infty} df' \int_{-\infty}^{+\infty} dt' \tilde{K}(f') e^{2\pi if'(t-t')} e^{-2\pi i ft'} = \int_{-\infty}^{+\infty} df' \tilde{K}(f') e^{2\pi if't} \int_{-\infty}^{+\infty} dt' e^{-2\pi i ft'} \delta(f+f')
\]

which upon integration yields

\[\tilde{K}(-f)e^{-2\pi if t}\]

and we may identify the eigenvalue \(\gamma_f = \tilde{K}(-f)\). The conventional application is to the spectral analysis of stationary time series. The statement of translation invariance is just the definition of stationarity. Then \(t\) is time and \(f\) is frequency.

3.9 Mechanical Systems

Consider a horizontal tight string with both ends clamped. Let the vertical displacement from equilibrium be \(\Psi(x,t)\) The string satisfies the one dimensional wave equation

\[\frac{1}{c^2} \frac{\partial^2}{\partial t^2} \Psi(x,t) = \frac{\partial^2}{\partial x^2} \Psi(x,t)\]

with boundary conditions

\[\Psi(0,t) = \Psi(L,t) = 0\]

and initial conditions

\[\Psi(x,0) = I(x)\]
\[\frac{\partial \Psi}{\partial t} \bigg|_{t=0} = V_0(x)\]

We solve the PDE by the method of separation of variables. Assume

\[\Psi(x,t) = X(x)T(t)\]

subbing in:

\[\frac{X''}{X} = \frac{1}{c^2} \frac{T''}{T} = -\lambda^2\]
CHAPTER 3. EMPIRICAL ORTHOGONAL FUNCTIONS

where for consistency (the LHS is a function of \( x \), the RHS is a function of \( t \)) the two ratios must be a constant, namely the separation constant \(-\lambda^2\). The solution for \( X(x) \) is

\[
X(x) = A \cos \lambda x + B \sin \lambda x
\]  

(3.126)

The only way to enforce the boundary conditions is to have

\[
A = 0, \quad \lambda = \frac{n\pi}{L}, \quad n = 1, 2, \ldots
\]  

(3.127)

whence

\[
X(x) = B_n \sin \left( \frac{n\pi}{L} x \right), \quad n = 1, 2, \ldots
\]  

(3.128)

Similarly,

\[
T(t) = C_n \cos \left( \frac{n\pi}{L} ct \right) + D_n \sin \left( \frac{n\pi}{L} ct \right)
\]  

(3.129)

And since if \( \Psi_1 \) and \( \Psi_2 \) are solutions so are \( a\Psi_1 + b\Psi_2 \), we can take the linear combinations

\[
\Psi(x, t) = \sum_{n=0}^{\infty} \left[ C_n \cos \left( \frac{n\pi}{L} ct \right) + D_n \sin \left( \frac{n\pi}{L} ct \right) \right] \sin \left( \frac{n\pi}{L} x \right)
\]  

(3.130)

The last is a general solution of the wave equation subject to the end point boundary conditions. The coefficients \( C_n \) and \( D_n \) can now be found by the initial conditions.

Now suppose the initial conditions were chosen from random initial conditions. I.e., suppose

\[
D_n, C_n \sim N(0, \sigma_n^2)
\]  

(3.131)

(Pluck the string randomly!). Note that \( C_n, D_n \) determine how much energy is in each normal mode. \((\propto C_n^2 + D_n^2)\).

Next consider the random field resulting. We can look at the string after an interval \( t_0 \), record our realization, restart the string with a new set of random \( C_n, D_n \); after \( t_0 \) we have another realization, etc.

Clearly,

\[
\Psi(x) = \sum_n F_n \sin \left( \frac{n\pi}{L} x \right)
\]  

(3.132)

and

\[
\langle F_n F_m \rangle = \lambda_n \delta_{mn}
\]  

(3.133)

Now these last are just the conditions that we have found for the Karhunen-Loeve expansion! The EOFs are

\[
e_n(x) = A_n \sin \left( \frac{n\pi}{L} x \right), \quad n = 1, \ldots
\]  

(3.134)
where the $A_n$ are normalization coefficients:

$$A_n = \sqrt{\frac{2}{L}} \quad (3.135)$$

### 3.9.1 Sampling errors for EOFs

### 3.9.2 Degenerate eigenvalues

**Exercise**

Observations have been taken at many points along the line $-1 \leq x \leq 1$. Each realization yields a straight (but different) line versus $x$ and they have been fitted to the form:

$$T(x) = A + Bx$$

where $A$ and $B$ are rv’s, $\sim N(0, \Sigma)$. In other words, each realization produces a new pair $(A, B)$, and they are correlated.

$$\Sigma = \begin{pmatrix} V_A & C \\ C & V_B \end{pmatrix}$$

1) What is $\langle T(x) \rangle$?

2) What is $\sigma_T^2(x) = \langle T(x)^2 \rangle$? Sketch it as a function of $x$.

3) What is the pdf of $T(x)$ in terms of $x, A, B$? (In the pdf please use $a$ instead of $A$, $b$ instead of $B$, etc., why?)

4) What is $K(x, x') = \langle T(x)T(x') \rangle$?

5) Set up the eigenvalue problem for the eigenfunctions (EOFs) $\psi_n(x)$.

6) Using the trial form $\psi_n(x) = \alpha_n + \beta_n x, \quad n = 1, 2$; find $\lambda_1, \lambda_2$ and $(\alpha_1, \beta_1), (\alpha_2, \beta_2)$; do not forget the normalization.

7) Express $T(x)$ in terms of the EOFs and new rv’s which are statistically independent.

8) Express the pdf of $T(x)$ in terms of the $\psi_1(x)$ and $\psi_2(x)$. 
Chapter 4

Time Series and Stochastic Processes

Consider a function $X(t)$ which is bounded. Now consider a collection of such functions (the ensemble). The individual members (realizations) are uncorrelated with each other. The ensemble average $\langle X(t) \rangle$ is given by

$$\langle X(t) \rangle = \lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} x(m)(t) \quad (4.1)$$

4.1 Stationarity

$$\langle X(t) \rangle = \langle X(t') \rangle \quad (4.2)$$

and

$$\langle X(t)X(t') \rangle = f(|t - t'|) \quad (4.3)$$

where $t - t'$ is the ‘lag’ and $f$ is a function.

4.2 Gaussian Process

Consider discrete times $t_1, t_2, \ldots$, Then the multivariate pdf $P(x(t_1), x(t_2), \ldots)$ is $G(0, \sigma_{ij})$. The covariance matrix need not be diagonal. The Gaussian assumption is often made although some important time series do not satisfy it, for example the telegraph signal or point measurements of rain.
CHAPTER 4. TIME SERIES AND STOCHASTIC PROCESSES

4.3 Discrete Time Series

We shall deal first with discrete time series

\[ X_1, X_2, X_3, \ldots, X_N \] (4.4)

4.3.1 White Noise

Definition:

\[ X_1, X_2, X_3, \ldots \] (4.5)

\[ X_j \text{ an } r.v. \in G(0, \sigma^2) \] (4.6)

\[ \langle X_n \rangle = 0 \] (4.7)

\[ \langle X_n X_m \rangle = \sigma^2 \delta_{nm} \] (4.8)

Each value in a white noise time series is uncorrelated with all past and future members. In estimating the average,

\[ \bar{X} = \frac{1}{N} \sum_{n=1}^{N} X_n, \] (4.9)

\[ \bar{X} \to \langle X \rangle \] (4.10)

4.4 Moving Average Process, MA₁

\[ Y_n = aZ_n + BZ_{n-1} \] (4.11)

where \( Z_n \) are gaussian white noise,

\[ \langle Z_n \rangle = 0, \] (4.12)

\[ \langle Z_n Z_m \rangle = \sigma_Z^2 \delta_{nm} \] (4.13)

Consider the product \( Y_n Y_m \)

\[ Y_n Y_m = a^2 Z_n Z_m + abZ_n Z_{m-1} + abZ_{n-1} Z_m + b^2 Z_{n-1} Z_{m-1} \] (4.14)

Taking the ensemble average

\[ \frac{\langle Y_n Y_m \rangle}{\sigma_Z^2} = (a^2 + b^2)\delta_{n,m} + ab\delta_{n,m-1} + ab\delta_{n-1,m} + b^2\delta_{n,m} \] (4.15)

Let \( m = n + l \), where \( l \) is the ‘lag’. Then

\[ \frac{\langle Y_n Y_{n+l} \rangle}{\sigma_Z^2} = (a^2 + b^2)\delta_{0,l} + ab(\delta_{l,1} + \delta_{l,-1}) \] (4.16)
Note first that the covariance $\langle Y_nY_{n+l} \rangle$ does not depend on $n$ since $Y_n$ is a stationary time series.

The quantity

$$\rho_l = \frac{\langle Y_nY_{n+l} \rangle}{\langle Y_n^2 \rangle} \quad (4.17)$$

is called the lagged correlation function. Note that by definition $\rho_1 = 1$. In general, $|\rho_l| \leq 1$ for $l > 0$.

### 4.5 Random Walk Process

Consider the process

$$Y_{n+1} - Y_n = \alpha Z_n, \quad Z_n \in G(0, \sigma_Z^2 \delta_{nm}) \quad (4.18)$$

What is $\rho_l$? We need to solve the last equation. To do this sum it from 0 to $N - 1$:

$$\sum_{n=0}^{N-1} (Y_{n+1} - Y_n) = \alpha \sum_{n=0}^{N-1} Z_n \quad (4.19)$$

All terms cancel except those on the ends:

$$Y_N - Y_0 = \alpha \sum_{n=0}^{N-1} Z_n \quad (4.20)$$

As an example take $Y_0 = 0$ (initial condition = the mean for all members of the ensemble). Compute the variance after $N$ steps.

$$\langle Y_N^2 \rangle = \alpha^2 \sum_{n,m=0}^{N-1} = \alpha^2 N \sigma_Z^2 \quad (4.21)$$

or

$$\sigma_{Y_N}^2 = \alpha^2 N \sigma_Z^2 \quad (4.22)$$

which grows as $N$ grows. The random walk the variance of which grows without bound is not really a stationary process. The random walk is important because it is a stochastic model of diffusion (Brownian Motion).

### 4.6 Autoregressive Process, AR$_1$

Next consider a modification of the RW process:

$$Y_{n+1} - Y_n = -\gamma Y_n + \alpha Z_n \quad (4.23)$$
CHAPTER 4. TIME SERIES AND STOCHASTIC PROCESSES

where the damping coefficient ($\gamma > 0$) provides a stabilizing feedback to prevent the variance of $Y_n$ from growing indefinitely. The continuous version of the last is the Brownian Motion problem in mechanics.

\[
m \frac{dv}{dt} = -bv + f(t)
\]

where $m$ is mass of the particle, $b =$ frictional damping coefficient, $v =$velocity, and $f(t) =$white noise forcing due to the random buffeting by molecules (originally solved by Einstein).

To solve our finite difference version, we write

\[
Y_{n+1} - \lambda Y_n = \alpha Z_n, \lambda = 1 - \gamma \leq 1
\]

Multiply through by $(1/\lambda)^{n+1}$:

\[
\frac{Y_{n+1}}{\lambda^{n+1}} - \frac{Y_n}{\lambda^n} = \frac{\alpha Z_n}{\lambda^{n+1}}
\]

Now sum from 0 to $N - 1$ and use the method used earlier:

\[
\frac{Y_N}{\lambda^N} - \frac{Y_0}{\lambda^0} = \frac{\alpha}{\lambda} \sum_{n=0}^{N-1} \frac{Z_n}{\lambda^n}
\]

Again if we take $Y_0 = 0$,

\[
Y_0 = 0
\]

\[
Y_1 = \alpha Z_0
\]

\[
Y_2 = \frac{\lambda^2 \alpha}{\lambda} \left( Z_0 + \frac{Z_1}{\lambda} \right)
\]

4.6.1 Computation of Lagged Covariance

\[
\langle Y_N Y_{N+l} \rangle = \alpha^2 \sum_{n=0}^{N-1} \sum_{m=0}^{N+l-1} \lambda^{N-1-n} \lambda^{N+l-1-m} \langle Z_n Z_m \rangle
\]

\[
= \alpha^2 \sigma_Z^2 \lambda^{2N} \sum_{n=0}^{N-1} \left( \frac{1}{\lambda^2} \right)
\]

\[
= \alpha^2 \sigma_Z^2 \lambda^{2N} \lambda^l \left( \frac{1 - \lambda^{-2N}}{1 - \lambda^{-2}} \right)
\]

\[
= \alpha^2 \sigma_Z^2 \lambda^l \frac{1 - \lambda^{2N}}{1 - \lambda^2}
\]
For $N$ very large this becomes

$$\langle Y_N Y_{N+t}\rangle = \left( \frac{\alpha^2 \sigma_Z^2}{1 - \lambda^2} \right) \lambda^t$$

(4.35)

$$\frac{\langle Y_N Y_{N+t}\rangle}{\langle Y_N Y_N \rangle} = \lambda^t$$

(4.36)

which is independent of $N$ (the series becomes stationary after a few [how many?] steps). Eventually knowledge of the initial condition $Y_0 = 0$ is forgotten(!) and the transient solution has died out.

We can now get the autocorrelation

$$\rho_l = \lambda^l, |\lambda| < 1$$

(4.37)

Integral time scale:

$$\tau = \sum_{l=0}^{\infty} \rho_l = \sum_{0}^{\infty} \lambda^l = \frac{1}{1 - \lambda}$$

(4.38)

$$= \frac{1}{\gamma} = \frac{1}{\text{damping coeff}}$$

(4.39)

$$\tau = \text{Correlation Interval}$$

(4.40)

This last is the characteristic time for correlation to decay.

### 4.6.2 Variance of the AR$_1$ Process

$$\langle Y_N^2 \rangle = \frac{\alpha^2 \sigma_Z^2}{1 - \lambda^2} 1 - \lambda^{2N}$$

(4.41)

After normalizing:

$$\frac{\langle Y_N^2 \rangle}{\langle Z^2 \rangle} = 1 - \lambda^{2N}$$

(4.42)

Note that in the damped case the variance is bounded and the process becomes stationary as $N \to \infty$.

### 4.7 Prediction in an AR$_1$ Process

Suppose $Y_0 \neq 0$, then we can write

$$Y_N = Y_0 \lambda^N + \alpha \lambda^{N-1} \sum_{0}^{N-1} \frac{Z_n}{\lambda^n}$$

(4.43)
Figure 4.1: The lagged autocorrelation function for an AR1 process after \( n \) lags (heavy dots). Standard deviation of \( Y_n \) is shown after \( n \) steps (heavy dots joined by dotted line. In both cases, \( \lambda = \rho_1 = 0.8, \tau = 5 \).

The ensemble average of all the trajectories passing through \( Y_0 \) at \( n = 0 \) is
\[
\langle Y_N \rangle = Y_0 \lambda^N, \text{ since } \langle Z_n \rangle = 0
\] (4.44)

The average trajectory returns to zero as \( \lambda^N \) (same as \( \rho_N \)). In atmospheric science we think of three kinds of statistical prediction: 1) persistence \((Y_N = Y_0)\), 2) climatology \((Y_N = 0)\) and 3) damped persistence \((Y_N = \lambda^N \text{ (Lorenz)}\).

Next we need an estimate of the error in a damped persistence (AR1) forecast. That is \( V(Y_N) \). We can readily compute
\[
V(Y_N) = (1 - \lambda^{2N})V(Y_\infty)
\] (4.45)

4.8 Linear Prediction Models

Suppose we want to predict \( Y_1 \) based upon a knowledge of \( Y_0 \). We also assume that we can write the predictor of \( Y_1 \) as a linear function of the previous value \( Y_0 \):
\[
\hat{Y}_1 = aY_0
\] (4.46)
where $a$ is a value to be chosen so as to make the prediction best in the least squares error sense. We can write for the mean square error

$$e^2(a) = \langle (Y_n - Y_{n-1})^2 \rangle$$  \hspace{1cm} (4.47)

Then

$$\frac{\partial e^2}{\partial a} = (Y_{n-1}(Y_n - aY_{n-1})) = 0$$  \hspace{1cm} (4.48)

or

$$a = \frac{\langle Y_{n-1}Y_n \rangle}{\langle Y^2_{n-1} \rangle}$$  \hspace{1cm} (4.49)

$$= \rho_1$$  \hspace{1cm} (4.50)

This leads us to the prediction

$$\hat{Y}_1 = \rho_1 Y_0$$  \hspace{1cm} (4.51)

$$\hat{Y}_2 = \rho_2 Y_1 \sim \rho_1^2 Y_0$$  \hspace{1cm} (4.52)

Hence, if only one value from the past is given we arrive at the damped persistence as our model. The general problem of finding $Y_n$ when many past values are known boils down to finding $E(Y_n|Y_{n-1}, Y_{n-2}, \ldots, Y_{n-k})$, i.e., the mean of the conditional probability distribution. A linear model is

$$\hat{Y}_n = a_1 Y_{n-1} + a_2 Y_{n-2} + \cdots + a_k Y_{n-k} + Z_n$$  \hspace{1cm} (4.53)

where $Z_n$ is white noise. This is the AR$_k$ model. We need to find the coefficients $a_1, a_2, \ldots, a_k$. We need to find estimates of the $a_i$ by least squares. We use the form

$$\sum_j s_{ij}\hat{a}_j = s_{yi}$$  \hspace{1cm} (4.54)

where we used

$$s_{ij} = \frac{S_{ij}}{p} = \sum_k (Y_i(k) - \bar{Y}_i)(Y_j(k) - \bar{Y}_j)/p$$  \hspace{1cm} (4.55)

where overbar indicates sample mean, $k$ is an index indicating realization number, and $p$ is the number of degrees of freedom in the sum of squares. In this case $s_{ij}$ is the estimate of Cov($Y_{n-i}, Y_{n-j}$). Dividing by $\langle Y^2_n \rangle$ we get autocorrelations

$$\sum_j \hat{\rho}(|j-i|)\hat{a}_j = \hat{\rho}(n-i), i = 1, \cdots, k$$  \hspace{1cm} (4.56)

The hat is used on $\rho(m)$ since we are dealing with sample estimates of autocorrelations. The last system of equations are known as the Yule-Walker equations.
4.9 Sampling Errors and Autocovariance

Consider the discrete process $X_n, n = 1, \cdots, N$. We estimate the mean by

$$\bar{X} = \frac{1}{N} \sum_{1}^{N} X_n$$  \hspace{1cm} (4.57)

The lag covariance by

$$\hat{\gamma}_k = c_k = \frac{1}{N} \sum_{k+1}^{N} (X_n - \bar{X})(X_{n-k} - \bar{X})$$ \hspace{1cm} (4.58)

Now $c_k$ is a random variable due to sampling. We can find its variance by standard methods

$$\text{Cov}(c_k, c_l) = \frac{1}{N} \sum_{r=-\infty}^{\infty} (\gamma_r \gamma_{r+l-k} + \gamma_{r+l} \gamma_{r-k})$$ \hspace{1cm} (4.59)

Similarly for the lag correlation coefficient

$$\text{Cov}(r_k, r_{k-s}) \approx \frac{1}{N} \sum_{-k+s}^{k} \rho_n \rho_{n-s}$$ \hspace{1cm} (4.60)

and setting $s = 0$:

$$\text{Var}(r_k) \approx \frac{1}{N} \sum_{-k}^{k} \rho_n^2$$ \hspace{1cm} (4.61)

An important point: sample correlations are correlated with themselves at different lags. Just because the function $\hat{r}_l$ is a nice smooth function does not mean it is statistically significant. Note that there will be nonzero estimates of $\gamma_k$ even when it is theoretically zero. Suppose we know that $\rho_l = 0$ for $l > L$. Then we can use Bartlett’s formula to see how $r_l$ will fluctuate beyond $L$.

$$\sigma_{r_k} = \sqrt{\frac{1}{N} \sum_{-L}^{L} \rho_n^2}, \quad k > L$$ \hspace{1cm} (4.62)

to get the 95% CI use $\pm 2\sigma_{r_k}$. 

4.10 Climatic Noise Due to Sampling

Suppose we take a month long average of a meteorological time series, say temperature, $X_n$. If $\bar{X}$ is the 30 day average, we certainly expect some fluctuation just due to sampling. In fact if $X_n$ were white noise, we would expect

$$\sigma^2_{\bar{X}} = \frac{\sigma^2_X}{N}, N = 30.$$  \hspace{1cm} (4.63)

Now suppose the true mean $\langle X \rangle$ is really different from one year say to the next. Then $\Delta \langle X \rangle$ had better be large or at least comparable to $\sigma_{\bar{X}}$ or we cannot make effective use of our knowledge about $\Delta \langle X \rangle$, since it will be buried in the 'noise' of sampling a single thirty day average. Leith calls this 'climatic noise'.


The problem is made complicated by the fact that the $X_n$ are not white noise but are correlated more like an AR1 process with characteristic time $(1/\gamma = 1/1-\lambda)$ of usually about 3 days (midlatitude temperature fluctuations at a point). The problem then is that there are much fewer than 30 independent samples in a month, making $\sigma^2_{\bar{X}}$ much larger! Next consider calculating the correction.

We have

$$\sigma^2_{\bar{X}} = \langle \bar{X}^2 \rangle = \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \langle X_n X_m \rangle$$  \hspace{1cm} (4.64)

At this point we could use the data to find the empirical form of $\rho_l$, or we could adopt the AR1 model (see Jones, R. H. 1975: Estimating the variance of time averages. J. Appl. Meteorol. 14, 159-163.) We take

$$\rho_l = \lambda^{|l|}, \ |\lambda| < 1$$  \hspace{1cm} (4.65)

Then

$$\frac{\langle \bar{X}^2 \rangle}{\sigma^2_{\bar{X}}} = \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \lambda^{|n-m|}$$  \hspace{1cm} (4.66)

After much algebra and neglecting terms of order $N^{-2}$ we get

$$\frac{\langle \bar{X}^2 \rangle}{\sigma^2_{\bar{X}}} = \frac{1}{N} \frac{1 - \lambda}{1 + \lambda}$$  \hspace{1cm} (4.67)
We can simplify further using $\gamma = 1 - \lambda$, $\tau = 1/\gamma$, $1 + \lambda = 2 - 1/\tau$, leading us to the approximation

$$N_{eff} \approx \frac{N}{2\tau}$$

(4.68)

which says that the interval between independent samples is twice the autocorrelation time.

There have been many studies establishing the level of climatic noise. The observed variability is of course always larger than the climatic noise and the difference represents “potential predictability” of climate. (see Madden, 1981: Atmospheric Variability and prediction. JGR, 170, 9817.)
Chapter 5

Four Fourier Transforms

We often have to deal with time series in the frequency domain. This can be a very useful undertaking, since the information content though equivalent may be more clearly revealing to the investigator. There are four types of time series that are encountered in geosciences.

1. Continuous time on the infinite domain: \(-\infty < t < \infty\).

2. Continuous time on a finite domain: \(-T/2 \leq t \leq T/2\).

3. Discrete time \(\ldots, t_{-1}, t_0, t_1, \ldots; t_{n+1} - t_n = \Delta t\), on the infinite domain: \(-\infty < t_n < \infty\).

4. Discrete time \(t_n\) on a finite domain \(-(N-1)/2 < n < (N-1)/2\); (here we take \(N\) to be an odd integer to keep the notation from being too cluttered).

5.1 Continuous Time on Infinite Domain

The classical Fourier transform of a function \(g(t)\) is defined as

\[
\tilde{g}(f) \equiv \int_{-\infty}^{\infty} g(t) e^{2\pi if t} \, dt.
\]  

(5.1)

The Dirac delta function \(\delta(t - t')\) has already been introduced. It satisfies the important identity

\[
\int_{-\infty}^{\infty} e^{2\pi if(t-t')} \, df = \delta(t - t')
\]  

(5.2)
Multiplying (5.1) on the left by $e^{-2\pi if't}$ and integrating wrt $f$

$$\int_{-\infty}^{\infty} \tilde{g}(f)e^{-2\pi if't} \, df = \int_{-\infty}^{\infty} \delta(t - t')g(t) \, dt$$

(5.3)

where we used the delta identity (5.2). We then have

$$g(t) = \int_{-\infty}^{\infty} \tilde{g}(f)e^{-2\pi ift} \, df$$

(5.4)

which is known as the Inverse Fourier Transform. The formulas (5.1) and (5.4) are known as a Fourier Transform pair. Given a function $g(t)$, we can find its FT: $\tilde{g}(f)$; conversely, given $\tilde{g}(f)$, we can recover the original function by integration.

### 5.1.1 Box Function

Consider the box function

$$B(t) = \begin{cases} 
0, & t < -L/2 \\
\frac{1}{L}, & -L/2 \leq t \leq L/2 \\
0, & t > L/2 
\end{cases}$$

(5.5)

The FT of the box function is

$$\tilde{B}(f) = \frac{\sin \pi Lf}{\pi Lf}$$

(5.6)

which is unity at $f = 0$ and has zeros at $f = n2/L$, $n = \pm 1, \pm 2, \cdots$

### 5.1.2 Gaussian

Consider the function

$$g(t) \equiv \frac{1}{\sqrt{2\pi\sigma}} \exp \left(-\frac{t^2}{2\sigma^2}\right)$$

(5.7)

The FT is

$$\tilde{g}(f) = \exp \left(-2\sigma^24\pi^2f^2\right)$$

(5.8)

We see that when $g(t)$ is “skinny”, $\tilde{g}(f)$ is “fat”. That is

$$\Delta t \cdot \Delta(2\pi f) = \frac{1}{2}$$

(5.9)

which is known as the uncertainty principle. Note that it also holds for the boxcar function.
5.1. Continuous Time on Infinite Domain

5.1.3 Lorentzian

Consider finally the Lorentzian function $g(t)$

$$g(t) = e^{-|t|/\tau} \quad (5.10)$$

We find that

$$\tilde{g}(f) = \frac{2\tau}{1 + 4\pi^2 f^2 \tau^2} \quad (5.11)$$

5.1.4 Convolution

Consider the convolution of a function $g(t)$ with the function $h(t)$ defined as

$$g \star h \equiv \int_{-\infty}^{\infty} g(t) h(\tau - t) \, dt \quad (5.12)$$

Note that $g \star h = h \star g$. This form is frequently encountered in the case of filtering or smoothing with “running” averages. For example, if we had an irregular function $g(t)$, we might want to make a running average of width $L$ on it with the boxcar function:

$$\bar{g}(t) = \int_{-\infty}^{\infty} g(t') B(t - t') \, dt' \quad (5.13)$$

which is a convolution of $g$ with $B$.

Consider the FT of a convolution. First insert the FTs of $g$ and $h$ into (5.12).

$$(g \star h)(t) = \int \int \tilde{g}(f) e^{-2\pi if't} \tilde{h}(f') e^{-2\pi if'(t' - t)} \, df \, df' \, dt' \quad (5.14)$$

$$= \int \int \delta(f + f') \tilde{g}(f) \tilde{h}(f') e^{2\pi if't} \, df \, df' \quad (5.15)$$

$$= \int \tilde{g}(f) \tilde{h}(f') e^{-2\pi if't} \, df \quad (5.16)$$

This last is the statement that the FT of $g \star h$ is just $\tilde{g} \tilde{h}$. In the case of the boxcar smoother, we have that the FTs multiply:

$$\tilde{\bar{g}}(f) = \frac{\sin \pi Lf}{\pi Lf} \cdot \tilde{g}(f) \quad (5.17)$$
### 5.1.5 Integral Identities

We can establish an important relationship between a function and its FT. Consider the integral over the entire time domain

\[ \int_{-\infty}^{\infty} g(t) \, dt = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{g}(f) e^{-2\pi if t} \, dt \, df \]  

(5.18)

\[ = \int_{-\infty}^{\infty} \delta(f) \tilde{g}(f) \, df \]  

(5.19)

\[ = \tilde{g}(0) \]  

(5.20)

Hence the integral over the entire time domain is the zero frequency component of the function.

Next consider the integral of the square of the (real) function.

\[ \int_{-\infty}^{\infty} g^2(t) \, dt = \int \int \tilde{g}(f) e^{-2\pi if t} \tilde{g}(f') * e^{+2\pi if' t} \, dt \, df \]  

(5.21)

\[ = \int_{-\infty}^{\infty} \vert \tilde{g}(f) \vert^2 \, df \]  

(5.22)

(5.23)

This last is known as Parseval’s identity. It states that the total “power” is the same in each domain.

### 5.2 Continuous Time, Finite Domain

In this case we examine functions on a finite temporal interval: \(-T/2 \leq t \leq T/2\). The Fourier Series representation on this domain is

\[ g(t) = \sum_{n=-\infty}^{\infty} \tilde{g}_n e^{-2\pi int/T} \]  

(5.24)

The (complex) Fourier coefficients are given by

\[ \tilde{g}_n = \frac{1}{\sqrt{T}} \int_{-T/2}^{T/2} g(t) e^{2\pi int/T} \, dt \]  

(5.25)

We are using the basis set

\[ \phi_n(t) = \frac{1}{\sqrt{T}} e^{2\pi int/T} \]  

(5.26)
5.2. CONTINUOUS TIME, FINITE DOMAIN

These basis functions have the orthonormality property
\[ \int_{-T/2}^{T/2} \phi_n(t) \phi_m^*(t) \, dt = \delta_{nm} \] (5.27)
and the completeness relation
\[ \sum_{n=-\infty}^{\infty} \phi_n(t) \phi_n^*(t') = \delta(t-t') \] (5.28)
Parseval’s relation assumes the form:
\[ \int_{-T/2}^{T/2} g^2(t) \, dt = \sum_n \sum_m \tilde{g}_n \tilde{g}_m \int_{-T/2}^{T/2} \phi_n(t) \phi_m^*(t) \, dt \]
\[ = \sum_{n=-\infty}^{\infty} |\tilde{g}_n|^2 \] (5.29)
\[ = \sum_{n=-\infty}^{\infty} |\tilde{g}_n|^2 \] (5.30)
The function as represented by the Fourier Series is periodic on the interval \(-T/2 \leq t \leq T/2\).
\[ g(t) = g(t + nT), \quad n = 0, \pm 1, \ldots \] (5.31)
The original function of course may not be periodic but its representation is.

5.2.1 Random fields

Suppose the statistics of the field are rotationally invariant on the circle. That is, \( \langle g(t) \rangle = 0 \), (we take the mean to be zero without loss of generality since if it is constant it could have been subtracted out.) Also assume that
\[ \langle g(t)g(t') \rangle = C(|t-t'|) \]
, where the covariance \( C(r) \) is some smooth function. Here \( t \) can be thought of as a point along the circumference of a circle, with \( T \) the total circumference. We may write
\[ \langle \tilde{g}_n \tilde{g}_m^* \rangle = \frac{1}{T} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \langle g(t)g(t') \rangle e^{2\pi i(nt-mt')/T} \, dt \, dt' \] (5.32)
\[ = \frac{1}{T} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \sum_l C_l e^{2\pi il(t-t')/T+2\pi i(nt-mt')/T} \, dt \, dt' \] (5.33)
\[ = \sum_l \delta_{ln} \delta_{lm} C_l \] (5.34)
\[ = C_n \delta_{nm} \] (5.35)
But we can also calculate

\[ C_n = \langle |g_n|^2 \rangle \]  

(5.36)

These last statements indicate that the complex random variables \( g_n \) are uncorrelated with each other except for the special case \( g_n \) and \( g_{-n}^* \).

If we have only one realization of the random function \( g(t) \), we can form the FT of it on the finite interval \((-T/2, T/2), g_n\). With this one realization we could form \( |g_n|^2 \), the periodogram, and use it as an estimate of the spectral density \( \langle |g_n|^2 \rangle = C_n \).

In the next chapter we can ask such questions as suppose we really have a process on the whole infinite interval, but just have data on the finite interval \((-T/2, T/2)\). How does our estimate of the spectrum compare with the corresponding spectrum from the infinite interval stationary process.

An important lesson about the finite interval process is that although the time is continuous, the frequency spectrum breaks into a countable (but infinite) set of discrete spectral “lines” which run from

\[ f_n = -\infty, \ldots, -2/T, -1/T, 0, 1/T, 2/T, \ldots \infty \]

Hence, the resolution in frequency is \( \Delta f = 1/T \). From the point of view of trying to infer something about the infinite interval process we see that the shortness of the record limits the resolution in the frequency domain. After all with a record of length \( T \) we should not expect to see frequencies lower than \( 1/T \).

### 5.3 Discrete Interval, Infinite Domain

In this case we are dealing with data collected at discrete intervals, but we have an infinitely long record. Consider the function

\[ g(t_n), t_n = -\infty, \ldots, -t_1, 0, t_1, t_2, \ldots, \infty \]

We take the data to be collected at equal intervals, \( \Delta t \), so that \( t_n = n\Delta t \).

The appropriate representation is (inverse transform)

\[ g(t_n) = \frac{1}{f_c} \int_{-f_c}^{f_c} \tilde{g}(f)e^{-\pi i f t_n / f_c} \, df \]  

(5.37)

and the FT is

\[ \tilde{g}(f) = \sum_{n=-\infty}^{\infty} g(t_n)e^{\pi i n f / f_c} \]  

(5.38)

and

\[ f_c = \frac{1}{2\Delta t} \]  

(5.39)
5.4. DISCRETE INTERVAL; FINITE DOMAIN

is called the “Nyquist frequency”. The Parceval relation is

$$\sum_{n=-\infty}^{\infty} g(t_n)^2 = \int_{-f_c}^{f_c} |\tilde{g}(f)|^2 df$$  \hspace{1cm} (5.40)

This time we have a continuum of frequencies but they are in the range $-f_c \leq f \leq f_c$. In the case of a stochastic process, we might consider that the data were taken on finite intervals, but we have infinitely long realizations. We have the interesting situation that the highest resolvable frequency is twice the Nyquist frequency. Certainly we do not expect to know much about the very high frequency behavior of a function if we only can examine it at discrete intervals.

5.4 Discrete Interval; Finite Domain

Finally we come to the very interesting case which actually comes up most often in practice. We have data at discrete intervals

$$-(N-1)/2, \ldots, -1, 0, 1, \ldots, (N-1)/2.$$  \hspace{1cm} (where we have taken the number of data points as before to be odd since to constantly repeat the slightly different formulas for even $N$ would be unnecessarily cumbersome.) Since $N$ is finite, we also have a finite domain for the process.

A few useful identities are

$$\sum_{l=0}^{L-1} r^l = \frac{1 - r^L}{1 - r}$$  \hspace{1cm} (5.41)

For $N$ odd, $M = (N-1)/2$

$$\sum_{n=-M}^{M} r^n = \frac{r^{-M+1} - r^M}{1 - r} = -r^{1/2} \left( \frac{r^M - r^{-M}}{r^{1/2} - r^{-1/2}} \right)$$  \hspace{1cm} (5.42)

The Fourier pairs are

$$\hat{x}_s = \frac{1}{N} \sum_{n=-M}^{M} x_n e^{2\pi i f_s n}, \quad f_s = \frac{s}{N}$$  \hspace{1cm} (5.43)

$$x_n = \sum_{s=-M}^{M} \hat{x}_s e^{-2\pi i f_s n}, \quad -(N-1)/2N \leq f_s \leq (N-1)/2N$$  \hspace{1cm} (5.44)
The sequences are periodic in both $n$ and $s$:

\[
\begin{align*}
x_{n+N} &= x_n \\
\tilde{x}_s &= \tilde{x}_{s+N}
\end{align*}
\]

We also have that for $x_n$ real, $\tilde{x}_s = \tilde{x}_s^*$, and at zero frequency: $\tilde{x}_0 = \tilde{x}_0^*$.

The Parseval formula is

\[
\sum_{n=-M}^{M} x_n^2 = N \sum_{s=-M}^{M} |\tilde{x}_s|^2
\]

5.4.1 Sines and Cosines

The last case is encountered so often it is important to know its expression in the real representation which uses ordinary sines and cosines instead of complex exponentials. First the identities for $N$ odd:

\[
\begin{align*}
\sum_{s=-M}^{M} \cos(2\pi f_s n) \cos(2\pi f_s m) &= \frac{N}{2} \delta_{nm}, \ n > 0 \\
\sum_{s=-M}^{M} \sin(2\pi f_s n) \sin(2\pi f_s m) &= \frac{N}{2} (\delta_{nm} - \delta_{m0}) \\
\sum_{s=-M}^{M} \sin(2\pi f_s n) \cos(2\pi f_s m) &= 0
\end{align*}
\]

We write

\[
x_n = \sum_{s=-M}^{M} \{A_s \cos(2\pi f_s n) + B_s \sin(2\pi f_s n)\}
\]

and using the identities we can find the Fourier coefficients:

\[
\begin{align*}
A_s &= \frac{2}{N} \sum_{n=-M}^{M} x_n \cos(2\pi f_s n), \ s > 0, \\
B_s &= \frac{2}{N} \sum_{n=-M}^{M} x_n \sin(2\pi f_s n) \\
A_0 &= \frac{1}{N} \sum_{n=-M}^{M} x_n = \tilde{x}_n
\end{align*}
\]
The Parseval formula is

\[ \sum_{n=-M}^{M} x_n^2 = N A_0^2 + \frac{N}{2} \sum_{s=1}^{M} (A_s^2 + B_s^2) \]  

(5.56)
Chapter 6

Spectral Analysis of Time Series

6.1 Discrete Fourier Transforms

We have a record of \( N \) entries

\[ X_0, X_1, \ldots, X_{N-1} \quad (6.1) \]

Since our formulae depend upon whether \( N \) is odd or even let’s take \( N \) to be odd. We seek a Fourier representation

\[ X_n \rightarrow \tilde{X}_s \quad (6.2) \]

with corresponding angular frequency

\[ \omega_s = \frac{2\pi}{N} s, \quad s = -M, -1, 0, 1, 2, \ldots M \quad (6.3) \]

or regular cycles/time frequency \( f_s = s/N \). Write

\[ X_n = \sum_{s=-M}^{M} \tilde{X}_s e^{i\omega_s n} \quad (6.4) \]

Multiply by \( e^{-i\omega_u n} \), sum \( n = 0 \) to \( N - 1 \)

\[ \sum_{n=0}^{N-1} e^{-i\omega_u n} X_n = \sum_{s=-M}^{M} \tilde{x}_s \sum_{n=0}^{N-1} e^{i(n(\omega_s - \omega_u))} \quad (6.5) \]
CHAPTER 6. SPECTRAL ANALYSIS OF TIME SERIES

\[ \sum_{s=-M}^{M} \hat{X}_s \left( \frac{1 - e^{iN(\omega_s - \omega_u)}}{1 - e^{i(\omega_s - \omega_u)}} \right) = N \delta_{su} \]  
(6.6)

\[ \hat{X}_s = \frac{1}{N} \sum_{n=0}^{N-1} X_n e^{-i\omega_s n} \]  
(6.8)

The formulae (6.4) and (6.8) are the discrete Fourier Transform and its inverse. We yet to calculate the value of \( M \).

Multiply (6.8) through by \( e^{i\omega_s m} \), sum on \( s \)

\[ \sum_{s=-M}^{M} e^{i\omega_s m} = \frac{1}{N} \sum_{n=0}^{N-1} X_n \sum_{s} e^{i\omega_s (m-n)} \]  
(6.9)

\[ = \frac{1}{N} \sum_{n=0}^{N-1} X_n \cdot \left( \frac{e^{-(M+1)i\omega_{m-n}} - e^{(M+1)i\omega_{m-n}}}{1 - e^{2i\omega_m}} \right) \]  
(6.10)

\[ = (\#\ of\ terms) \delta_{mn} \]  
(6.11)

To make it work we must take \( \#\ of\ terms = N \), therefore

\[ M = \frac{N - 1}{2}, \ N \text{odd} \]  
(6.12)

Example: \( N=5, M=2 \). Hence, sum contains \( s = -2, -1, 0, 1, 2 \) or 5 terms, the same number as in the original time series. Finally

\[ X_n = \sum_{-(N-1)/2}^{(N-1)/2} \hat{X}_s e^{i\omega_s n} \]  
(6.13)

6.2 Symmetric Time Interval, \( \infty \) Interval

Take \( N \) odd as before; \( N \) is the total number of values of \( X_n \). Then our formulas go through

\[ X_n = \sum_{s=-M}^{M} \hat{X}_s e^{2\pi inf_s} \]  
(6.14)

\[ \hat{X}_s = \frac{1}{N} \sum_{n=-(N-1)/2}^{(N-1)/2} X_n e^{-2\pi inf_s} \]  
(6.15)

with

\[ f_s = \frac{s}{N}, \ \omega_s = 2\pi f_s \]  
(6.16)

Note that \( X_n \) is periodic: \( X_{n+N} = X_n \) and also \( \hat{X}_s \) is periodic: \( \hat{X}_{s+N} = \hat{X}_s \).
6.2. **SYMMETRIC TIME INTERVAL, \( \infty \) INTERVAL**

### 6.2.1 Passing to the \( \infty \) interval

In the continuous time domain think of \( t = n; \delta t = \Delta n(= 1); T = N \):

\[
\tilde{X}_s = \frac{1}{T} \int_{-T/2}^{T/2} X(t)e^{-2\pi if_j t} dt
\]

\[
T\dot{X}_s \equiv \dot{X}_f \rightarrow \int_{-\infty}^{\infty} X(t)e^{-2\pi if_j t} dt
\]

which is the Fourier Integral representation. The inverse is

\[
x(t) = \int_{-\infty}^{\infty} \dot{X}_f e^{2\pi if_j t} df
\]

The formulas (6.15) and (6.14) are the discrete Fourier transformation formulae for data starting from \( n = 0 \) and running up to \( N - 1 \), \( N \) terms in all.

### 6.2.2 Discrete Fourier Amplitudes

\( X_n \) is real.

\[
\tilde{X}_s = \sum_n X_n e^{-i\omega_s n}
\]

\[
\tilde{X}_s^* = \sum_n X_n e^{+i\omega_s n}
\]

\[
\tilde{X}_s^* = \sum_n X_n e^{-i\omegaS n}
\]

\[
\tilde{X}_s = \dot{X}_s
\]

### 6.2.3 Variances

Let \( \langle X_n \rangle = 0 \), \( X_n \) stationary.

\[
\tilde{X}_s = \frac{1}{N} \sum_n X_n e^{-i\omega_s n}
\]

look at the covariance:

\[
\tilde{X}_s \tilde{X}_{s+u}^* = \frac{1}{N^2} \sum_{n,m} X_n X_m e^{-i\omega_s n + i\omega_{s+u} m}
\]

\[
= \frac{1}{N^2} \sum_{n,m} X_n X_m e^{i\omega_s (m-n) + i\omega_u m}
\]
CHAPTER 6. SPECTRAL ANALYSIS OF TIME SERIES

First look at the expected value

\[
\langle \tilde{X}_s \tilde{X}_s^* \rangle = \frac{1}{N^2} \sum_{n,m} \gamma_{n-m} e^{i\omega_s (m-n) + i\omega_u m}
\]  

(6.27)

As an example, consider \(X_n\) to be white noise. Then

\[
\gamma_{n-m} = \sigma^2 \delta_{nm}
\]  

(6.28)

Then

\[
\langle \tilde{X}_s \tilde{X}_{s+u}^* \rangle = \frac{1}{N^2} \sum_n \sigma^2 e^{i\omega_u n}
\]  

(6.29)

\[
= \frac{\sigma^2}{N^2} \left( \frac{1 - e^{i\omega_u N}}{1 - e^{i\omega_u}} \right) e^{-i\omega_u N/2}, \quad \omega_u = \frac{2\pi u}{N}
\]  

(6.30)

For \(u = 0\) this is undefined, but just go back to the previous formula:

\[
\langle |\tilde{X}_s|^2 \rangle = \frac{\sigma^2}{N} = S(f_s)
\]  

(6.31)

This last \(S(f_s)\) is called the power spectrum, \(f_s = \frac{\omega_s}{2\pi}\). Note that the power spectrum is independent of \(s\) and equivalently \(f_s\) or \(\omega_s\). [picture of a series of equal sized vertical bars, on the \(s\) axis running from \(-M \leq s \leq M\)] The power spectrum for white noise is flat. We may also plot \(S(f)\) (note that the subscript is dropped); in this case is is called the spectral density. [picture the vertical bars spaced \(\Delta f = \frac{1}{N}\) apart. The plot runs from \(-\frac{1}{2} \leq f \leq \frac{1}{2}\). There are a total of \(N\) lines] The highest frequency “1/2” is called the Nyquist frequency.

6.3 Non-White Variance

The lag covariance may be expanded:

\[
\gamma_{n-m} = \sum_{t=-M}^M \tilde{\gamma}_t e^{i\omega_t (n-m)}
\]  

(6.32)

Then

\[
\langle \tilde{X}_s \tilde{X}_{s+u}^* \rangle = \frac{1}{N^2} \sum_n \sum_m \langle X_n X_m \rangle e^{2\pi i f_s n + 2\pi i f_u m}
\]  

(6.33)

\[
= \frac{1}{N^2} \sum_{t=-M}^M \tilde{\gamma}_t \sum_{n,m} e^{2\pi i (f_t - f_s) n} e^{2\pi i (f_{u}^* - f_t) m}
\]  

(6.34)
6.4. AR\(_1\) SPECTRUM

\[
\hat{\gamma}_t = \frac{1}{n^2} \sum_{t=-M}^{M} \hat{\gamma}_t N^2 \delta_{t,s} \delta_{t,s'}
\]

or finally

\[
\hat{\gamma}_t = S(f_s)
\]

In other words, \(\hat{\gamma}_t\), the Fourier transform of the lag covariance \(\gamma_{n-m}\) is the spectrum.

Note that \(\hat{X}_s\) and \(\hat{X}_{s'}^*\) are uncorrelated unless \(s = s'\). Roughly speaking this means that each spectral line is uncorrelated with the other spectral lines: Hence

\[
X_n = \sum_s \hat{X}_s e^{2\pi if_s n}
\]

is like a Karhunen-Loeve expansion. Looking at the inverse FT shows that \(\hat{X}_s\) is an rv and it is uncorrelated with its neighbors.

One way to compute an estimate of the spectrum is to form the FT of the lag covariance.

\[
\hat{\gamma}_s = S(f_s) = \sum_{l=-M}^{M} \gamma_{|l|} e^{-2\pi il f_s}
\]

If we know \(\gamma_l\) we can then compute \(S(f)\). When the subscript \(s\) is dropped we call \(S(f)\) the spectral density.

6.4 AR\(_1\) Spectrum

\[
\gamma_l = \sigma^2 \lambda^{|l|}
\]

Then

\[
\hat{\gamma}_s = \sigma^2 \sum_{l=-(N-1)/2}^{(N-1)/2} \lambda^{|l|} e^{-2\pi il f_s}
\]

Note first that \(\hat{\gamma}_s\) is real. For algebraic simplicity, compute the \(T \to \infty\) limit

\[
T \hat{\gamma}_f = \sigma^2 \int_{-\infty}^{\infty} \lambda^{|t|} e^{-2\pi if t} dt
\]

\[
S(f) = \sigma^2 \int_{0}^{\infty} e^{-|t| \ln t - 2\pi f t} dt
\]
chapter 6. spectral analysis of time series

\[ + \sigma^2 \int_0^\infty e^{-|t| \ln t + 2\pi if t} dt \]  
\[ = \frac{\sigma^2}{-\ln \lambda - 2\pi f} + \frac{\sigma^2}{-\ln \lambda + 2\pi f} \]  
\[ = \frac{\sigma^2 \ln(1/\lambda)}{(\ln \lambda)^2 + (2\pi f)^2} (> 0) \]

[picture of the spectrum]

This spectrum, peaked at low frequencies, is characteristic of geophysical phenomena, since they frequently arise from “Brownian Motion” types of processes. The concentration of power in the low frequencies is usually called a “red noise” spectrum in analogy with color spectra.

6.5 Some Fourier Transforms

Consider the box function

\[ g_n = \begin{cases} g_0 & -\frac{(l-1)}{2} < n < \frac{(l-1)}{2} \\ 0 & \text{otherwise} \end{cases} \]  

whose total width is \( l \).

\[ \tilde{g}_s = \frac{1}{N} \sum_{n=-(N-1)/2}^{(N-1)/2} g_n e^{-2\pi if_s n} \]  
\[ = \frac{g_0}{N} \sum_{n=-(l-1)/2}^{(l-1)/2} e^{-2\pi if_s n} \]  
\[ = \frac{g_0}{N} \sin(\pi f_s) \sin(\pi f_s) \]

Then

\[ |\tilde{g}_s|^2 = S(f) \]

[picture of the function with lobes etc. ] Note the width in \( f \) space is \( 2/l \), but the width in time space was \( l \). We have the identity

\[ w_f \cdot w_{time} = 2 \]

6.5.1 Pulse

Suppose \( l = 1 \), the width of the box. Then

\[ S(f) \rightarrow \text{constant} \]
6.6. SUMS OF UNCORRELATED TIME SERIES

In the continuous case this corresponds to

\[ g_n = g(t) = g_0 \delta(t) \]  \hspace{1cm} (6.51)

Then

\[ \tilde{g}_f = g_0 \int_{-\infty}^{\infty} \delta(t) e^{-2\pi i ft} \, dt \]  \hspace{1cm} (6.52)
\[ = g_0 \]  \hspace{1cm} (6.53)

6.5.2 Pure Tone

Suppose \( g(t) = A \sin 2\pi f_0 t \). Then

\[ g(t) = \frac{A}{2i} \left[ e^{2\pi i f_0 t} - e^{-2\pi i f_0 t} \right] \]  \hspace{1cm} (6.54)

Then

\[ \tilde{g}_f = \frac{A}{2i} \int_{-\infty}^{\infty} \left[ e^{(2\pi i f_0 t - 2\pi i ft)} - e^{-(2\pi i f_0 t + 2\pi ft)} \right] \, dt \]  \hspace{1cm} (6.55)
\[ = \frac{A}{2i} [\delta(f - f_0) - \delta(f + f_0)] \]  \hspace{1cm} (6.56)

The power is concentrated at \( f = \pm f_0 \). A problem arises because \([\delta(f)]^2\) is not defined. This is traceable back to the fact that we are dealing with a deterministic signal, not a stochastic one. In fact, the power at \( \pm f_0 \) depends on the length of the record. The longer the record, the more prominent will be the peak.

Consider the discrete case

\[ g_n = \Re \left( g_0 e^{2\pi i f_A n} \right) \]  \hspace{1cm} (6.57)

and suppose that \( f_A = f_s \) for one of the discrete values, of \( s \). Then

\[ \tilde{g}_s = \frac{g_0}{2N} \sum_n \left[ e^{2\pi i n(f_A - f_s)} + e^{2\pi i n(f_A + f_s)} \right] \]  \hspace{1cm} (6.58)
\[ = \frac{g_0}{2N} [\delta_{A,s} + \delta_{A,-s}] \]  \hspace{1cm} (6.59)

6.6 Sums of Uncorrelated Time Series

Suppose \( X_n, Y_n \) are random processes such that

\[ \langle X_n \rangle = \langle Y_n \rangle = 0 \]  \hspace{1cm} (6.60)
\[ \langle X_n Y_n \rangle = 0 \]  \hspace{1cm} (6.61)
Let
\[ U_n = \alpha X_n + \beta Y_n \] (6.62)
then
\[ C(l) = \langle U_n U_{n+l} \rangle \] (6.63)
\[ = \alpha^2 C_x(l) + \beta^2 C_y(l) \] (6.64)
Since
\[ \tilde{\gamma}_s = S(f) \] (6.65)
we have
\[ S_u(f) = \alpha^2 S_x(f) + \beta^2 S_y(f) \] (6.66)
The spectral densities can be added.

The most common example is of the form
\[ Y_n = \underbrace{X_n}_{\text{Noise such as AR}_1} + \underbrace{g_n}_{\text{Signal such as tone}} \] (6.67)

We frequently encounter this problem of trying to estimate a deterministic signal buried in noise or natural variability. One issue is the resolution of the spectrum \( \Delta f = 1/N \). We can increase our resolution by making \( N \) larger.

### 6.7 Errors in Spectral Estimation

#### 6.7.1 Aliasing

If we are examining a continuous process by sampling at finite intervals \( \Delta t \), we can find fictitious power that should be given to higher, unresolved frequencies. [pic of wave with wavelength L above one with 2L. Sample at L, see how the higher freq aliases. Aliasing then leads us to think there is power in our spectrum at certain frequencies when it really is due to frequencies that are too high to resolve.]

#### 6.7.2 Leakage

This is the opposite problem. Suppose \( N \) is too small so that we cannot resolve some low frequency actually in the infinitely long time series. It will show up.
6.7. ERRORS IN SPECTRAL ESTIMATION

6.7.3 Sampling Errors

Sampling errors are normally an important problem in estimating spectra in geosciences. First note that

\[ \hat{X}_s = \frac{1}{N} \sum_{n=-M}^{M} X_n e^{-2\pi i n f_s} \]  \hfill (6.68)

is an rv and the spectral line intensity

\[ |\hat{X}_s|^2 = \frac{1}{N^2} \sum_{n,m} X_n X_m e^{-2\pi i f_s (n-m)} \]  \hfill (6.69)

is also an rv. We wish to find its pdf. Note that we assume the \( X_n \) are a stationary Gaussian time series.

As before we treat white noise first. Let

\[ R_s = \frac{1}{N} \sum X_n \cos(\omega_s n) \]  \hfill (6.70)

\[ I_s = \frac{1}{N} \sum X_n \sin(\omega_s n) \]  \hfill (6.71)

It is easy (?) to show that

\[ \text{Cov}(R_s, I_{s'}) = 0 \]  \hfill (6.72)

Also note that \( R_s \) = weighted sum of gaussian rv's. Hence, \( R_s \) and \( I_s \) are Gaussian (this will hold even if the \( X_n \) are not G if \( N \) is large. Then the \( R_s^2 \) and \( I_s^2 \) are (except for normalization) \( \chi^2 \) variables. It follows that

\[ |\hat{X}_s|^2 = R_s^2 + I_s^2 \]  \hfill (6.73)

\[ \propto \chi^2_{[2]} \]  \hfill (6.74)

and

\[ \langle R_s^2 \rangle = \frac{\sigma^2}{N} \]  \hfill (6.75)

Hence,

\[ |\hat{X}_s|^2 = \chi^2_{[2]} \frac{\sigma^2}{N} \]  \hfill (6.76)

We can now learn about the sampling errors because we know about the pdf of \( \chi^2_{[2]} \).

\[ \langle \chi^2_{[\nu]} \rangle = \nu = \# \text{dof} \]  \hfill (6.77)

\[ \text{Var}(\chi^2_{[\nu]}) = 2\nu \]  \hfill (6.78)
We see that in this case for a single spectral line we have \( \nu = 2 \); i.e., each spectral line has 2 dof. Then if \( \hat{S}(f_s) \) is a spectral estimator at the frequency \( f_s \) from a sample segment of length \( N \), we know that it is an \( \text{rv} \in \chi^2_2 \), and thus

\[
\text{Var}(\hat{S}(f_s)) = \frac{\sigma^4}{N^2} \cdot 2 \cdot 2
\]

or

\[
\text{Standard Error} = 2 \frac{\sigma^2}{N} = \langle \hat{S}(f_s) \rangle
\]

The standard error in estimating \( S(f_s) \) is just \( \langle \hat{S}(f_s) \rangle \) itself! Note that unlike the situations we encountered earlier, there is no \( 1/\sqrt{N} \) around to make things better if we work harder! [pic of line spectrum with error bars]. We could find a 95% confidence region with \( \chi^2_2 \) tables. [pic of a typical white noise estimated spectrum (periodogram). Note that the frequency resolution is \( \Delta f = 1/N \).

### 6.8 Using Lagged Covariance

We showed that potentially convenient way of computing the spectrum is to Fourier Transform the lag covariance function. Suppose we estimate the lag covariance \( \hat{\gamma}_l \) from the available record of \( N \) entries according to the earlier formula. Clearly, we cannot estimate \( \gamma \) for lags \( l \) comparable to \( N \). We probably should not use values of \( l > N/10 \) in practice. Hence, if we try to use the FT on \( \hat{\gamma}_l \) we have to make some assumption about its behavior for the larger values of \( l \). The usual technique is to “window” \( \hat{\gamma}_l \) by multiplying by a function \( w_l \)

\[
\hat{\gamma}_l' = \hat{\gamma}_l w_l
\]

A typical \( w_l \) might be the “boxcar” or rectangular function

\[
w_l = \begin{cases} 
1, & |l| \leq \frac{L-1}{2} \\
0, & |l| > \frac{L-1}{2}
\end{cases}
\]

This is equivalent to setting \( \hat{\gamma}_l \) equal to zero beyond \( |l| > (L - 1)/2 \). Note that the Fourier Transform of a product like \( \hat{\gamma}_l w_l \) is the convolution

\[
\mathcal{F}(\hat{\gamma}_l w_l) = \hat{\gamma}_s * \hat{w}_s
\]

\[
= \sum_u \hat{\gamma}_s u \hat{w}_{u-s}
\]
6.9 Spectral Smoothing

A popular method of estimating spectra is to compute the raw periodogram by Fourier transforming the original timeseries segment and squaring. Usually before this is done the first and last 10% of the times series are tapered smoothly to zero. This will remove some spurious high frequency power due to the sharp cutoff of the time series. [pic of time series and a tapering function]. The modified periodogram is produced by squaring the modified Fourier coefficients. The band width remains $\Delta f_s = 1/N$.

Now a simple moving average is applied to the spectral estimate. Let the moving average have bandwidth $b$. Then the dof for this bandwidth is simply $2N/b$.


The effect of “convolving” $\tilde{\gamma}_s$ with $\tilde{w}_{u-s}$ is like applying a running average of $\tilde{\gamma}_s$ with the squared weights $\tilde{w}_{u-s}$. Hence, the window procedure is a way of smoothing $\tilde{\gamma}_s$. [pic of a raw periodogram, a smoothing filter with lobes, then the product. ] The problem is that $\tilde{w}_{u-s}$ has side lobes (the first is about 1/5 the size of the central maximum and negative). Various “lag windows” other than the “boxcar” have been proposed (one famous one by Parzen) which have better properties but the principle is the same.


By averaging $N/L$ spectral lines together as above we increase the bandwidth from $1/N$ to $1/L$ ($L$ is the fraction of the total time series length given to the filter). In so doing, we increase the number of dof from 2 to $2N/L$. For a white spectrum we can return to the expression (6.74) to obtain

$$\tilde{\gamma}'_s = \chi^2 [2N/L] \frac{\sigma^2}{2N}$$

(6.85)

Then

$$\langle \tilde{\gamma}'_s \rangle = \frac{2N}{L} \frac{\sigma^2}{2N} = \text{smoothed spectrum}$$

(6.86)

We also find that

$$\text{Var}(\tilde{\gamma}'_s) = \frac{2}{4} \frac{\sigma^4}{N^2} \frac{2N}{L \text{dof}}$$

(6.87)

We should be reminded that this applies to a case for which a smoothing operator of width $1/L$ has been applied ($1/L =$ bandwidth = $b$). The standard
error of the spectrum is

\[ SE_{S(f)} = \frac{\sigma^2}{2N} \cdot 2 \sqrt{\frac{N}{L}} = \langle \hat{\gamma}'_s \rangle \cdot \sqrt{\frac{L}{N}} \]  \hspace{1cm} (6.88)

\[ = \langle \hat{\gamma}'_s \rangle \sqrt{\frac{1}{Nb}} \]  \hspace{1cm} (6.89)

It is clear from the last formula, that by increasing the bandwidth, \( b \), we can reduce the sampling variability. This is the basic trade off: bandwidth versus sampling error.

Since \( SE_{S(f)} \propto \langle \hat{\gamma}'_s \rangle \) it is usually convenient to express the spectrum on log paper. \([\text{pic of log S vers } f \text{ showing error bars and their invariance on this graph}].\)

\[ \text{width of the 95\% CI} \approx 4SE_{S(f)} = \frac{4}{\sqrt{Nb}} \langle \hat{\gamma}'_s \rangle \]  \hspace{1cm} (6.90)

Hence, on the log scale, \( \log \frac{4}{\sqrt{Nb}} \approx C + \log \langle \hat{\gamma}'_s \rangle \) and the same error bar (CI) applies to all levels of \( \langle \hat{\gamma}'_s \rangle \).

Strictly speaking our analysis only applies to white noise, but we can apply it approximately for other smooth spectra.

### 6.10 Bartlett Approach

Suppose we take our time series with \( N \) entries and divide it into \( Q \) segments of length \( N/Q \). We could compute the periodogram for each segment and average the \( Q \) periodograms. In this way we will get a \( 1/\sqrt{Q} \) kind of improvement in our estimate of the spectrum at \( f \). Notice, however, that we now have a spectral resolution (elementary bandwidth) of \( Q/N \) omstead of \( 1/N \) for the original whole series. \([\text{pic of the spectrum with the coarser resolution}].\) A detailed examination shows this to be equivalent to applying a triangular filter to the original data

\[ w_l = \begin{cases} (1 - |l|/L), & |l| \leq L \\ 0, & |l| > L \end{cases} \]  \hspace{1cm} (6.91)

### 6.11 Linear Systems

Consider the system

\[ \frac{dT}{dt} + \frac{T}{\tau} = g\delta(t) \]  \hspace{1cm} (6.92)
where $\delta(t)$ is the Dirac delta (impulse) function and $\tau$ is a relaxation time. Take the continuous FT:

$$-i2\pi f \hat{T}_f + \frac{\hat{T}_f}{\tau} = g$$

(6.93)

$$\hat{T}_f = \frac{g\tau}{1 - 2\pi i \tau f}$$

(6.94)

In frequency space terms the “flat” input $g$ is transformed by a filter $1/(1 - i\omega\tau)$. In terms of variance

$$|\hat{T}_f|^2 = \frac{1}{1 + (2\pi\tau)^2 f^2} \cdot g^2$$

(6.95)

[pic of the transform flat to lorentzian] Suppose the input is stochastic

$$\frac{dT}{dt} + \frac{T}{\tau} = g z(t)$$

(6.96)

Then we have

$$|\hat{T}_f|^2 = \frac{1}{1 + (2\pi\tau)^2 f^2} \cdot |\hat{z}_f|^2$$

$$S_T(f) = \frac{|h(f)|^2}{S_z(f)}$$

(6.97)

(6.98)

If we know two of these quantities we can find the third. Clearly from the impulse-response function (6.94), the filter can be derived. When noise is the input we can use the impulse response function to compute the spectral output of the linear system. Filters of this type are called “low pass” filters since only the low frequency part of the input “survives” to be realized as output.

Just as $1/(1 - 2\pi i \tau f)$ is a low pass filter, we could equally well design a high pass follow by a low pass to form a “bandpass” filter. For a good example see

6.12 Tests for White Noise

6.12.1 Testing $\rho = 0$

One way to proceed is to note that for white noise $\rho_0 = 1; \rho_l = 0, l > 0$. And we can use (for lag $k$):

$$\text{Var}(r_k) \approx \frac{1}{N} \sum_{-k}^{k} \rho_n^2$$

(6.99)

where $r_k$ is our sample estimate of $\rho$. For white noise then we should have ($N =$ series length):

$$\text{Var}(r_k) \approx \frac{1}{N}, \text{ all } k$$

(6.100)

Similarly we can show

$$E(r_k) \approx \left(1 - \frac{|k|}{N}\right) \rho_k$$

(6.101)

and $r_k$ is normally distributed about zero, for $N$ large. [pic of sample $r_k$ with many $k$. Note that about 1 in 20 should exceed the 95% level of C]

6.12.2 Kolmogorov-Smirnov test

Another approach is to use the fact that the spectrum should be flat. Consider the sample estimate of the spectrum

$$\hat{S}(f) = S_0 G(f)$$

(6.102)

where if we average $G(f)$ over $f$ or over an ensemble, $\bar{G} = 1$. The deviations of $G(f)$ from 1 are due to sampling error assuming the null hypothesis of white noise.

The K-S test examines the cumulative variance

$$I(f_s) = \frac{2}{N - 1} \sum_{s'=1}^{s} G(f_{s'})$$

(6.103)

Under $H_0$, $I(f_s)$ should be a straight line. [pic of straight line of I vs $f_s$, showing slope 2]

The significance test consists of constructing a band

$$\frac{\pm \lambda}{\sqrt{N/2 - 1}}$$

(6.104)
about the theoretical line. For significance levels of .95 and .75, \( \lambda \) is equal to 1.36 and 1.02. [pic of the five lines, dots showing data].

According to Jenkins and Watts: “The interpretation of the 75% limits, for example, is that in one out of four plots, on the average, the maximum deviation from the theoretical line will lie outside the limits if in fact the process is white noise.”
Chapter 7

Multivariate Time Series

Most of this chapter comes from Jenkins and Watts (JW).

7.1 Bivariate Time Series

Consider now the two time series $X_1^n$ and $X_2^n$. These may be related to each other in a variety of ways. Before getting to examples, we need to establish definitions. We want our pair of series to be stationary: $E(X_{in}) = 0$. (removing the mean, of course) We can then examine the joint pdf $P(x_1^n, x_2^n)$. Often, we will take it to be Gaussian, but remember that important exceptions occur. The autocovariance functions are

$$\gamma_{11}(l) = E(X_{1,n}X_{1,n+l}) \quad (7.1)$$
$$\gamma_{22}(l) = E(X_{2,n}X_{2,n+l}) \quad (7.2)$$

The cross covariance functions are

$$\gamma_{12}(l) = E(X_{1,n}X_{2,n+l}) \quad (7.3)$$
$$\gamma_{21}(l) = E(X_{2,n}X_{1,n+l}) \quad (7.4)$$

Note the properties

$$\gamma_{ii}(0) = \sigma^2_i = \text{var}(X_{in}) \quad (7.5)$$
$$\gamma_{ii}(l) = \gamma_{ii}(-l) \quad (7.6)$$
and

$$\gamma_{12}(l) = \gamma_{21}(-l) \quad (7.7)$$

This last says that knowledge of $\gamma_{12}(l)$ for all $l$ is equivalent to knowing $\gamma_{21}(l)$. 

107
CHAPTER 7. MULTIVARIATE TIME SERIES

The cross correlation function is defined by

$$
\rho_{12}(l) = \frac{\gamma_{12}(l)}{\sigma_1 \sigma_2}
$$

(7.8)

It can be shown that $|\rho_{12}(l)| \leq 1$.

An example is shown in the curve of $\gamma_{12}(l)$ with bump at about $l = 2$. Interpretation: $X_1$ and $X_2$ are correlated well when $X_2$ is evaluated two units later than $X_1$. We might say that $X_1$ leads $X_2$ and perhaps causes changes in $X_2$.

7.1.1 Analytical examples

Consider a simple example:

$$
X_{2n} = hX_{1n} + Z_n
$$

(7.9)

where $h$ is a constant. We take $Z_n$ to be white noise and assume $E(X_{1n}) = E(X_{2n}) = 0$. Then

$$
\gamma_{12}(l) = \gamma_{11}(l), \quad \text{assume } E(X_1Z) = 0
$$

(7.10)

and

$$
\gamma_{12}(0) = \gamma_{11}(0)
$$

(7.11)

(noise does not enter), but when we look at

$$
\gamma_{22}(0) = h^2\gamma_{11}(0) + \gamma_{zz}(0)
$$

(7.12)

Suppose $X_1$ is white noise, then

$$
\gamma_{11}(l) = \sigma_1^2 \delta_{l,0}
$$

(7.13)

and

$$
\gamma_{12}(l) = h\sigma_1^2 \delta_{l,0}
$$

(7.14)

More generally we can have the model

$$
X_{2n} = \sum_{l=-\infty}^{\infty} h_l X_{1,n+l} + Z_n
$$

(7.15)

with $Z_n$ white, $X_1$ uncorrelated with $Z_n$. Now multiply by $X_{2,n+l}$ and take expectations

$$
\gamma_{22}(l) = \sum_m \sum_{m'} h_m h_{m'} \langle X_{1,m+n} X_{1,m'+n+l} \rangle + \gamma_{ZZ}
$$

(7.16)

$$
= \sum_m \sum_{m'} h_m h_{m'} \gamma_{11}(m - m' - l) + \gamma_{ZZ}
$$

(7.17)
7.2. BIVARIATE ARMA

and similarly

\[ \gamma_{12}(l) = \sum_{m} h_m (X_{1,n}X_{1,n+l+m}) + 0 \] (7.18)

\[ = \sum_{m} h_m \gamma_{11}(l + m) \] (7.19)

hence, the \( \gamma_{22}(l) \) and \( \gamma_{12}(l) \) can be found if \( \gamma_{11}(l) \) is known. the factor \( h_l \)

is a digital filter.

7.1.2 Mixed inputs: the bivariate linear case

Consider the system

\[ X_{1n} = h_{11}Z_{1n} + h_{12}Z_{2n} \] (7.20)

\[ X_{1n} = h_{21}Z_{1n} + h_{22}Z_{2n} \] (7.21)

or in matrix form

\[ \mathbf{\tilde{X}}_n = \mathbf{H}\mathbf{\tilde{Z}}_n \] (7.22)

where \( Z_{1n} \) and \( Z_{2n} \) are uncorrelated white noise with variances \( \sigma_1^2 \) and \( \sigma_2^2 \).

\[ \gamma_{12}(0) = \sigma_1^2 h_{11} + h_{22} h_{12}^2 \sigma_2^2 \] (7.23)

\[ \gamma_{12}(l) = 0, \ l > 0 \] (7.24)

7.2 Bivariate ARMA

First consider the moving average process (MA):

\[ X_{1n} = \sum_{l} \beta_{1l} Z_{1,n+l} + \sum_{l} \beta_{2l} Z_{2,n+l} \] (7.25)

\[ X_{2n} = \sum_{l} \beta_{3l} Z_{1,n+l} + \sum_{l} \beta_{4l} Z_{2,n+l} \] (7.26)

Now the AR\( k \) model:

\[ X_{1n} = \sum_{l=1}^{k} \alpha_{11,l} X_{1,n-l} + \sum_{l=1}^{k} \alpha_{12,l} X_{2,n-l} + Z_{1n} \] (7.27)

\[ X_{2n} = \sum_{l=1}^{k} \alpha_{21,l} X_{1,n-l} + \sum_{l=1}^{k} \alpha_{22,l} X_{2,n-l} + Z_{2n} \] (7.28)
The $Z_{in}$ may be uncorrelated but an interesting exception is when they are correlated simultaneously

$$\langle Z_{in}Z_{jm} \rangle = \sigma_{ij}\delta_{nm} \quad (7.29)$$

Another interesting case is the continuous time AR$_1$:

$$\frac{d}{dt} \left( \begin{array}{c} X_1(t) \\ X_2(t) \end{array} \right) = \left( \begin{array}{cc} h_{11} & h_{12} \\ h_{21} & h_{22} \end{array} \right) \left( \begin{array}{c} X_1(t) \\ X_2(t) \end{array} \right) + \left( \begin{array}{c} Z_1(t) \\ Z_2(t) \end{array} \right) \quad (7.30)$$

and obviously one can mix AR and MA.

### 7.3 Cross Spectra

Consider the estimator $\hat{\gamma}_{12}(l)$ for $\gamma_{12}(l)$:

$$\hat{\gamma}_{12}(l) = \begin{cases} \frac{1}{N} \sum_{-M}^{M-l} X_{1,n}X_{2,n+l}, & 0 \leq l \leq N \\ \frac{1}{N} \sum_{-M+l}^{M} X_{1,n}X_{2,n+l}, & -N \leq l \leq 0 \end{cases} \quad (7.31)$$

This estimator is biased but tends to $\gamma_{12}$ as $N \to \infty$. The estimator $\hat{\gamma}_{12}(l)$ has the same undesirable property as $\hat{\gamma}_{11}(l)$, namely, neighboring estimates $\hat{\gamma}_{12}(l)$ and $\hat{\gamma}_{12}(l')$ for $l \approx l'$ are highly correlated. We take from Jenkins and Watts:

$$\text{Cov}(\hat{\gamma}_{12}(k), \hat{\gamma}_{12}(l)) \approx \frac{1}{N} \sum_{r=-\infty}^{\infty} [\gamma_{11}(r)\gamma_{22}(r + l - k) + \gamma_{12}(r + l)\gamma_{21}(r - k)] \quad (7.32)$$

This can be used to estimate the effect [pic of $\hat{\gamma}_{12}$ vs $l$ highlighting a small band between $l$ and $k$.]

#### 7.3.1 Effect of autocorrelation

Consider the interesting case of $X_1$ and $X_2$ uncorrelated, then $\gamma_{12} = 0$. Then

$$\text{Cov}(\hat{\gamma}_{12}(k), \hat{\gamma}_{12}(l)) \approx \frac{1}{N} \sum_{r=-\infty}^{\infty} \gamma_{11}(r)\gamma_{22}(r + l - k) \quad (7.33)$$

As an example take both to be AR$_1$:

$$\gamma_{11}(l) = \sigma_1^2|l| \quad (7.34)$$

and

$$\gamma_{22}(l) = \sigma_2^2|l| \quad (7.35)$$
Then when \( l = k \), we get

\[
\text{Var}(\hat{\gamma}_{12}(l)) \approx \frac{\sigma_1^2 \sigma_2^2}{N} \left( \frac{1 + \lambda_1 \lambda_2}{1 - \lambda_1 \lambda_2} \right) \quad (7.37)
\]

Now suppose \( \lambda_1, \lambda_2 > 0 \) (the usual case), we see that the factor in parentheses amplifies the factor \( \frac{\sigma_1^2 \sigma_2^2}{N} \), which is the white noise value. The sampling errors in estimating \( \hat{\gamma}_{12}(l) \) will be especially large then if \( \lambda_1 \) and \( \lambda_2 \) are close to unity (long autocorrelation times!)

### 7.4 Cross Spectra

We have two time series \( X_{1n}, X_{2n} \). We can look at their FTs:

\[
\hat{X}_{is} = \frac{1}{N} \sum_{n=-M}^{M} X_{in} e^{-2\pi ifs} \quad (7.38)
\]

\[
i = 1, 2; \quad 2M + 1 = N, \quad f_s = s/N \quad (7.39)
\]

In turn each can be written

\[
X_{in} = \sum_{s=-M}^{M} \hat{X}_{is} e^{2\pi ifs}, \quad \hat{X}_{is} = \hat{X}_{i,-s} \quad (7.40)
\]

The \( \hat{X}_{is} \) are random, complex variates for \(-M \leq s \leq M\). If the two time series are strongly correlated at some frequency, \( f_s \), then we expect

\[
\text{Cov}(\hat{X}_{1s}, \hat{X}_{2s}^*) \neq 0 \quad (7.41)
\]

It is natural to define the cross-spectrum as

\[
\hat{\gamma}_{12}(f_s) = \langle \hat{X}_{1s} \hat{X}_{2s}^* \rangle \quad (7.42)
\]

### 7.4.1 Example

Suppose each time series consists of an oscillation with frequency \( f_{s'} \):

\[
X_{jn} = \frac{A_j}{2} \left( e^{2\pi ifs_{n}} + e^{-2\pi ifs_{n}} \right) + \frac{B_j}{2i} \left( e^{2\pi ifs_{n}} - e^{-2\pi ifs_{n}} \right) \quad (7.43)
\]

This is just a superimposed sine and cosine.
In this case

\[ \tilde{X}_{js'} = \frac{1}{2} (A_j + B_j) \]  

(7.44)

\[ \tilde{X}_{j,-s'} = \frac{1}{2} (A_j - B_j) \]  

(7.45)

[note the symmetry \( \tilde{X}_{js'}^* = \tilde{X}_{j,-s'} \), since \( X_n \) is real]

For a stochastic process \( A_j, B_j \) would be real rv’s with mean zero. In this case then

\[ \tilde{\gamma}_{jj}(fs) = \frac{1}{4} ( (A_j^2 + B_j^2 ) ) \delta_{ss'} \]  

(7.46)

[pic of \( \tilde{\gamma}_{jj} \) vs \( s \) with a spike at \( s' \)]

If \( A_1 \) and \( A_2 \) are correlated and \( B_1 \) and \( B_2 \) are correlated we will get a cross-spectrum:

\[ \tilde{\gamma}_{12}(fs) = \frac{1}{4} ( (A_1 + B_1)(A_2 - B_2) ) \delta_{ss'} \]  

(7.47)

\[ = \frac{1}{4} \{ (A_1 A_2) + (B_1 B_2) + (A_1 B_2) - (A_2 B_1) \} \]  

(7.48)

Note that the cross spectrum is complex.

In general we may write

\[ \tilde{X}_{js} = A_{js} e^{i F_j(s)} \]  

(7.49)

where \( A_{js} \geq 0 \) is the modulus of \( \tilde{X}_{js} \) and \( F_j(s) \) is the phase of \( \tilde{X}_{js} \). Both are real rv’s. Now the sample cross spectrum can be written

\[ \hat{\tilde{\gamma}}_{12}(fs) = A_{1s} A_{2s} e^{i(F_2(s) - F_1(s))} \]  

(7.50)

and

\[ F_{12}(fs) \equiv F_2(s) - F_1(s) \]  

(7.51)

is called the sample phase spectrum while

\[ A_{12}(fs) \equiv A_{1s} A_{2s}, \text{ note } A_{12}(f) = A_{21}(-f) \]  

(7.52)

is called the sample amplitude spectrum Clearly the phase spectrum tells us what the phase lag at frequency \( f_s \) is between \( X_1 \) and \( X_2 \). When the two are uncorrelated it is undefined.
7.4.2 Co- and quadrature spectra

We could write
\[ \tilde{\gamma}_{12}(f_s) = L_{12}(f_s) - iQ_{12}(f_s) \] (7.53)
where
\[ L_{12}(f_s) = A_{12} \cos F_{12}(f_s) \] (7.54)
\[ Q_{12}(f_s) = -A_{12} \sin F_{12}(f_s) \] (7.55)

Similarly
\[ A_{12}^2 = L^2 + Q^2 \] (7.56)
and
\[ F_{12} = -\arctan \left( \frac{Q}{F} \right) \] (7.58)

$L$ is an even function of $f$, $Q$ is odd. $L_{12} \equiv$ co-spectrum measures the in phase components; $Q_{12} \equiv$ quadrature spectrum measures the out of phase components.

7.4.3 Example from JW

For two waves:
\[ X_1 = A_1 \cos(2\pi f_0 t - \phi_1) \] (7.59)
\[ X_2 = A_1 \sin(2\pi f_0 t - \phi_2) \] (7.60)

Then
\[ L_{12} = \frac{A_1 A_2}{4} \cos(\phi_1 - \phi_2)[\delta(f + f_0) + \delta(f - f_0)] \] (7.61)
\[ Q_{12} = \frac{A_1 A_2}{4} \sin(\phi_1 - \phi_2)[\delta(f + f_0) + \delta(f - f_0)] \] (7.62)

7.5 Covariance and Cross Spectrum

Again from JW we find that the relation is the same as before for auto spectra
\[ \tilde{\gamma}_{12}(f_s) = \frac{1}{N} \sum_{l=-M}^{M} \gamma_{12}(l)e^{-2\pi if_s l} \] (7.63)

Similarly
\[ \gamma_{12}(l) = \sum_{l=-M}^{M} \tilde{\gamma}_{12}(l)e^{2\pi if_s l} \] (7.64)
7.5.1 Cross-spectra examples (JW)

Bivariate Equivalent of white noise:
\[ X_2[n] = \beta_1 X_1[n] + Z_2[n], \quad X_1[n] = Z_1[n] \]  \hspace{1cm} (7.65)
and \( Z_{in} \) = white noise. Then
\[ \gamma_{12}(k) = \beta_1 \sigma_1^2 \delta_{k0} \]  \hspace{1cm} (7.66)
Then
\[ \tilde{\gamma}_{12}(f_s) = \beta_1 \sigma_1^2 \]  \hspace{1cm} (7.67)
The Cross-Amplitude spectrum:
\[ A_{12}(f_s) = \beta_1 \sigma_1^2, \quad F_{12} = 0 \]  \hspace{1cm} (7.68)
\[ L_{12} = \beta_1 \sigma_1^2, \quad Q_{12} = 0 \]  \hspace{1cm} (7.69)

JW: “(7.65) may be regarded as a fundamental model for cross spectra in
the same way that white noise is fundamental for univariate spectra”. 2)

Effect of delay
\[ X_2[n] = \beta_1 X_{1,n-d} + Z_2[n], \quad X_{1,n} = Z_{1n} \]  \hspace{1cm} (7.70)
then
\[ \gamma_{12}(k) = \beta_1 \sigma_1^2 \delta_{k,d} \]  \hspace{1cm} (7.71)
\[ \tilde{\gamma}(k) = \beta_1 \sigma_1^2 e^{-2\pi if_d} \]  \hspace{1cm} (7.72)
\[ A_{12}(f_k) = \beta_1 \sigma_1^2, \quad F_{12} = -2\pi f_k d \]  \hspace{1cm} (7.73)
\[ L_{12}(f_k) = \beta_1 \sigma_1^2 \cos(2\pi f_k d), \quad Q_{12}(f_k) = \beta_1 \sigma_1^2 \sin(2\pi f_k d) \]  \hspace{1cm} (7.74)

3) Now consider \( X_{2n} \) filtered from \( X_{1n} \)
\[ X_2[n] = \beta X_{1n} + \beta_2 X_{1,n-1} + Z_2[n] \]  \hspace{1cm} (7.76)
\[ X_1 = Z_{1n}, \quad Z_{in} = \text{white noise} \]  \hspace{1cm} (7.77)
Then
\[ \gamma_{12}(l) = \beta_1 \sigma_1^2 \delta_{0,l} + \beta_2 \sigma_1^2 \delta_{1,l} \]  \hspace{1cm} (7.78)
and then
\[ \tilde{\gamma}_{12}(f_s) = \sigma_1^2 (\beta_1 + \beta_2 e^{-2\pi if_s}) \]  \hspace{1cm} (7.79)

A final definition:
\[ \kappa_{12}^2 = \frac{|\tilde{\gamma}_{12}(f_s)|^2}{\tilde{\gamma}_{11} \tilde{\gamma}_{22}} \]  \hspace{1cm} (7.80)
is called the squared coherency spectrum. It is the analoge of cross-correlation
coefficient at each frequency. It is a nondimensional measure.
7.6 Estimating Cross-Spectra

White noise results: We have $\hat{A}_{i,s}$, $\hat{B}_{i,s}$ the Fourier Coefficients from series 1 and 2. (sine and cosine representation). We can show

$$\text{Var}(\hat{A}_i) = \text{Var}(\hat{A}_i) = \frac{N}{2} \sigma_i^2$$ (7.82)

$$\text{Cov}(\hat{A}_i, \hat{B}_i) = 0, \ i = 1, 2$$ (7.83)

Further if $Z_{1n}$ and $Z_{2n}$ are uncorrelated:

$$\text{Cov}(\hat{A}_1, \hat{A}_2) = 0$$ (7.84)

$$\text{Cov}(\hat{B}_1, \hat{B}_2) = 0$$ (7.85)

$$\text{Cov}(\hat{A}_1, \hat{B}_2) = 0$$ (7.86)

$$\text{Cov}(\hat{B}_1, \hat{A}_2) = 0$$ (7.87)

Also then

$$\langle \hat{L}_{12}(f_s) \rangle = 0$$ (7.88)

$$\text{Var}(\hat{L}_{12}(f_s)) = \frac{\sigma_1^2 \sigma_2^2}{2}$$ (7.89)

$$\langle \hat{Q}_{12}(f_s) \rangle = 0$$ (7.90)

$$\text{Var}(\hat{Q}_{12}(f_s)) = \frac{\sigma_1^2 \sigma_2^2}{2}$$ (7.91)

Nor for $\hat{A}_{12}^2$, use

$$\hat{\gamma}_{12}(f_s) = \hat{X}_1(f_s)\hat{X}_2(f_s)^*$$ (7.92)

then

$$\hat{A}_{12}^2 = |\hat{\gamma}_{12}|^2 = |\hat{X}_1(f_s)|^2 |\hat{X}_2(f_s)|^2$$ (7.93)

$$= \hat{\gamma}_{11}(f_s)\hat{\gamma}_{22}(f_s)$$ (7.94)

$$= \text{product of two } \chi^2 \text{ variates}$$ (7.95)

Then

$$\langle \hat{A}_{12}^2 \rangle = \langle \hat{\gamma}_{11}\hat{\gamma}_{22} \rangle$$ (7.96)

$$= 2 \cdot \hat{\gamma}_{11} \cdot 2 \cdot \hat{\gamma}_{22} = \frac{\sigma_1^2 \sigma_2^2}{N^2}$$ (7.97)

similarly

$$\text{Var}(\hat{A}_{12}^2) = 3 \cdot \frac{\sigma_1^4 \sigma_2^4}{N^2}$$ (7.98)
for a single series the 3 is missing.

Similarly it can be shown that the phase spectral estimator \( \hat{F}_{12}(f_s) \) is uniformly distributed on \(-\pi\) to \(\pi\).

### 7.6.1 Correlation test for two series

1. First convert the series to white noise. e.g. if \( X_{1n} \) is AR1, \( X_{1,n+1} = \lambda X_{1n} + Z_1 \), then filter it to obtain \( X'_{1n} = X_{1,n+1} - \hat{\lambda} X_{1n} \), where \( \hat{\lambda} \) is your estimate of \( \lambda \). Now \( X'_{1n} \) is a white noise series.

2. Now we want a test like the K-S test, i.e., an integral of the cross-spectrum. We choose
   \[
   \hat{L}_{12}(f_s) = \Re(\hat{\gamma}_{12})
   \]  
   (7.101)

Use the integrated form:
   \[
   \hat{J}_{12} = \frac{1}{N} \frac{2N^2}{\sigma_1 \sigma_2} \sum_{s=0}^{s'} \hat{L}_{12}(f_s)
   \]  
   (7.102)

While JW do not give rigorous criteria (!!!) this might give an idea about correlation. [pic of curves]

3. The phase spectrum \( F_{12} \) is supposed to be uniform hence the K-S test can be applied to it [pic with straight line \( \pm \lambda/\sqrt{N/2} \). For CL of 0.95, \( \lambda = 1.36 \); for CL of .75, \( \lambda = 1.02 \).

#### 7.7 Smoothed Estimators

In the univariate case we had
   \[
   \hat{\gamma}'(f_s) = \frac{1}{N} \sum_{l=-M}^{M} \hat{\gamma}'_l w_l \cos(2\pi f_s l)
   \]  
   (7.103)

where \( w_l \) is the so called ‘lag window’. We actually found that this leads to a convolution

   \[
   \hat{\gamma}'(f_s) = \hat{\gamma}'(f_s) \ast \hat{w}(f_s)
   \]  
   (7.104)

   \[
   = \sum_{s' \in \mathbb{Z}} \hat{w}(f - f_{s'}) \hat{\gamma}'(f_s)
   \]  
   (7.105)

In other words the lag window Fourier transforms into a frequency domain averager: \( \hat{w}(f - f_{s'}) \). [pic of box car transforming to a sin/x ] From the
7.7. SMOOTHED ESTIMATORS

last equation we get:

\[
E(\hat{\gamma}'_s) = \sum_{s'} \hat{w}(f - f_{s'}) E(\hat{\gamma}) \sigma^2 / N
\]

(7.106)

\[
= \frac{\sigma^2}{N} \sum_{s'} \hat{w}(f - f_{s'}) \text{boxcar}\n
\]

(7.107)

Now we can compute \( \text{Var}(\hat{\gamma}'_s) \), the variance of the smoothed estimator

\[
\text{Var}(\hat{\gamma}'_s) = \sum_s \sum_{s'} \hat{w}(s - s') \hat{w}(s - s'') \text{Cov}(\hat{\gamma}'_s, \hat{\gamma}'_{s''}) \delta_{s',s''} \text{Var}(\hat{\gamma}_s)
\]

(7.108)

Then for white noise we can factor out \( \text{Var}(\hat{\gamma}_s) \):

\[
\frac{\text{Var}(\hat{\gamma}'_s)}{\text{Var}(\hat{\gamma}_s)} = \sum_{s'} [\hat{w}(s - s')]^2
\]

(7.109)

\[
= \frac{\text{variance of smoothed}}{\text{variance of unsmoothed}}
\]

(7.110)

We can compute the RHS of the last equation for the box car filter:

\[
w_l = \begin{cases} 
0, & l > |(L - 1)|/2 \\
1, & l \leq |(L - 1)|/2
\end{cases}
\]

(7.111)

Use

\[
\hat{w}_s = \frac{1}{N} \sum_l w(l) e^{2\pi i f_s l}
\]

(7.112)

then

\[
\sum_s |\hat{w}|^2 = \frac{1}{N^2} \sum_{l,l'} w(l) w(l') \sum_s e^{-2\pi i f_s (l - l')} N \delta_{s,s'}
\]

(7.113)

or

\[
\sum_s |\hat{w}|^2 = \frac{1}{N} \sum_l [w(l)]^2 \text{ (Parseval’s formula)}
\]

(7.114)

(7.115)

(7.116)

Now for the boxcar,

\[
\sum_l [w(l)]^2 = L
\]

(7.117)
Hence for the boxcar:
\[
\frac{\text{Var}(\tilde{\gamma}_{s'})}{\text{Var}(\tilde{\gamma}_s)} = \frac{L}{N}
\]
(7.119)

It seems natural to renormalize the smoothed estimator
\[
\Gamma_s \equiv \frac{\sum_{s'} \hat{w}(s-s')\tilde{\gamma}_{s'}}{\sum_{s'} \hat{w}(s')}
\]
(7.120)

Then for white noise
\[
\langle \Gamma_s \rangle = \tilde{\gamma}_s \rightarrow \frac{\sigma^2}{2N\chi^2_{[2]}}
\]
(7.121)

and
\[
\text{Var}(\Gamma_s) = \left( \sum_{s'} |\hat{w}(s-s')|^2 \right) \text{Var}(\tilde{\gamma}_s)
\]
(7.122)
\[
= \frac{1}{N} \left( \sum_l w_l^2 \right) \text{Var}(\tilde{\gamma}_s)
\]
(7.123)

and
\[
\frac{\sqrt{\text{Var}(\Gamma_s)}}{\Gamma_s} = \sqrt{\frac{\sum_l w_l^2}{N}} \cdot 1
\]
(7.124)
\[
= \sqrt{\frac{2}{\nu_{\text{eff}}}}
\]
(7.125)

or for the smoothed estimator \( \Gamma_s \):
\[
\nu_{\text{eff}} = \frac{2N}{\sum_l w_l^2}
\]
(7.126)

Note that for any variate proportional to \( \chi^2_{[\nu]} \), say \( f = c\chi^2_{[\nu]} \), \( \text{Var}(f) = c^2\nu, \langle f \rangle = c\nu \). Then follows
\[
\frac{\sqrt{\text{Var}(f)}}{\langle f \rangle} = \sqrt{\frac{2}{\nu}}
\]
(7.127)

We can use the formula for \( \nu_{\text{eff}} \) to compute a confidence interval for \( \Gamma_s \) by taking it to be proportional to a \( \chi^2_{[\nu_{\text{eff}}]} \) variate. The procedure also works for smooth spectra (compared to \( L \)) not just white. Finally, the band width can be defined in analogy to the box car
\[
b = \frac{1}{\sum_l w_l^2} = \text{bandwidth}
\]
(7.128)
\[
= \frac{1}{L} \text{ for boxcar window}
\]
(7.129)
7.8. SMOOTHING CROSS-SPECTRA

Note the important rule:

\[
\text{Variance}(\Gamma_s) \times \text{bandwidth} = \text{constant} \quad (7.130)
\]

### 7.7.1 Bartlett Window

We could have used the housetop window:

\[
w_l = \begin{cases} 
1 - \frac{2l}{L-1}, & |l| \leq L - 1 \\
0, & |l| > L - 1 
\end{cases} \quad (7.131)
\]

Then

\[
\tilde{w}_s = \frac{L}{2N} \left[ \frac{\sin \left( \frac{\pi f L}{2} \right)}{\left( \frac{\pi f L}{2} \right)} \right]^2 
\]

(7.132)

Then we can calculate:

\[
\text{Variance Ratio} \rightarrow \frac{0.667 L}{2 \frac{L}{N}}, \quad \sum w_l^2 \approx L/3 \quad (7.133)
\]

\[
\text{dof} \rightarrow 6 \frac{L}{N} \quad (7.134)
\]

\[
\text{bandwidth} \rightarrow 3 \frac{L}{N} \quad (7.135)
\]

all of these last are better than the boxcar.

### 7.8 Smoothing Cross-Spectra

Cross spectra have the same undesirable property as in the univariate case: The periodogram variance does not tend to zero as \( N \rightarrow \infty \) as one might have hoped.

The cure is the same as before. Apply a lag window to your estimator and you will get a smoothed cross-spectrum. From the properties of the window you can get the effective dof and construct confidence limits.

One new feature does arise. If one series is related to the other by a delay, [pic of a delay example], it would be foolish to apply a lag window to this case. Instead the remedy is to realign the time scales so that the delay is removed. Then you proceed as before. For details see JW.
Chapter 8

Space-Time EOFs

In a number of estimation problems it is convenient to represent a random field in space-time by its EOFs in both space and time. The simplest case is that for a stationary multivariate process. The process is to involve \( J \) variates or 'station time series', \( T_j(t), \ j = 1, \ldots, J; t = 1, \ldots, N \). The space-time lagged covariance is

\[
\mathcal{M}_{jj'}(t - t') = \langle T_j(t)T_{j'}(t - t') \rangle \quad (8.1)
\]

The dependence on the difference, \( t - t' \) is because of the stationarity in time. Because of it, we can write

\[
\tilde{\mathcal{M}}_{jj'}(s) = \frac{1}{N} \sum_{\tau=1}^{N} \mathcal{M}_{jj'}(\tau)e^{-i2\pi s\tau/N} \quad (8.2)
\]

\[
\mathcal{M}_{jj'}(\tau) = \sum_{s=-M}^{M} \tilde{\mathcal{M}}_{jj'}(s)e^{i2\pi \tau s/N} \quad (8.3)
\]

The quantity \( \tilde{\mathcal{M}}_{jj'}(s) \) proportional to the cross-spectrum between stations \( j \) and \( j' \). We need to show this.

The cross-spectrum can be defined as

\[
\tilde{\gamma}_{jj'}(s) = \langle \tilde{T}_j(s)\tilde{T}_{j'}^*(s) \rangle \quad (8.4)
\]

\[
= \frac{1}{N} \sum_{t=1}^{N} \frac{1}{N} \sum_{t'=1}^{N} \tilde{\mathcal{M}}_{jj'}(t - t')e^{-i2\pi s(t-t')/N} \quad (8.5)
\]

then

\[
\tilde{\gamma}_{jj'}(s) = \sum_{s'=-M}^{M} \frac{1}{N} \sum_{t=1}^{N} \frac{1}{N} \sum_{t'=1}^{N} \tilde{\mathcal{M}}_{jj'}(s')e^{i2\pi t(s-s')/N}e^{-i2\pi t'(s-s')/N}
\]
\[ M = \sum_{s'} M_{jj'}(s') \delta_{ss'} \delta_{ss'} \]  
\[ = \tilde{M}_{jj'}(s) \]  

What about the eigenvectors of the kernel \( \langle T_j(t)T_{j'}(t') \rangle \)?

\[ \sum_{j=1}^{J} \sum_{t'=1}^{N} M_{jj'}(t-t')\psi_n(j',t') = \lambda_n \psi_n(j, t) \]  

Let us try the substitution

\[ \psi_n(j, t) = \phi_{m,s}(j) e^{i2\pi st/N} \]  

where we have split \( n \) into two indices, \( m \) and \( s \) (a space mode and a frequency index). Now subbing the FT of \( M_{jj'}(t-t') \) we obtain

\[ \sum_{j=1}^{J} \sum_{t'=1}^{N} \sum_{s'=1}^{M} \tilde{M}_{jj'}(s') e^{i2\pi (t-t')s'/N} \phi_{m,s}(j) e^{-i2\pi (t-t')s/N} = \lambda_{m,s} \phi_{m,s}(j) \]  

we can now perform the sum on \( t' \):

\[ N \sum_{j'} \sum_{s'} \tilde{M}_{jj'}(s') e^{i2\pi (s'-s)/N} \phi_{m,s}(j') = \lambda_{m,s} \phi_{m,s}(j) \]  

Now we can sum on \( t \) and rearrange:

\[ N \sum_{j=1}^{J} \tilde{M}_{jj'}(s) \phi_{m,s}(j) = \lambda_{m,s} \phi_{m,s}(j) \]  

Now this last is the statement that we can represent the space time EOFs in the factored form (8.9). We need only find the eigenvectors of \( N \tilde{M}_{jj'}(s) \) for each frequency \( s \). Now the matrix \( N \tilde{M}_{jj'}(s) \) is a hermitian matrix; i.e., its transpose complex conjugated is itself. A hermitian matrix has the properties that its eigenvalues are real and since its diagonal elements are positive real the eigenvalues will be positive. The eigenvectors are in the complex \( J \) dimensional space and they may point in different directions for different frequencies \( s \). The eigenvectors will have the usual properties:

\[ \sum_{j=1}^{J} \phi_{m,s}^*(j) \phi_{m',s'}(j) = \delta_{s,s'} \delta_{m,m'} \]  

\[ \sum_{s,m} \phi_{m,s}^*(j) \phi_{m,s}(j') = \delta_{j,j'} \]
The space-time covariance kernel can now be written:

\[ \mathcal{M}_{jj'}(t - t') = \sum_{m,s} \phi_{s,m}(j)e^{-i2\pi st/N} \lambda_{m,s} \phi_{m,s}(j')e^{i2\pi st'/N} \]  

(8.15)
Chapter 9

EOF Examples

Here we consider two examples of EOFs generated from theoretical models. The first example is based upon a diffusion model whose solution is examined after a fixed time interval from a set of random initial conditions. The second example is for waves on a string which is continually being agitated by a mad hammerer.

9.1 Diffusion Equation

We consider the solutions of the diffusion equation:

\[ \frac{\partial \Psi(r,t)}{\partial t} = D \nabla^2 \Psi(r,t) \quad (9.1) \]

We work in a domain \( r = (x,y) \) such that \( 0 < x, y < 1 \) with the boundary conditions that the field is always zero on the boundaries:

\[ \begin{align*}
\Psi(0,y,t) &= 0 & (9.2) \\
\Psi(1,y,t) &= 0 & (9.3) \\
\Psi(x,0,t) &= 0 & (9.4) \\
\Psi(x,1,t) &= 0 & (9.5)
\end{align*} \]

A physically realizable example of such a situation is a square block whose edges are in contact with a reservoir held at 0°, the field \( \Psi(x,y,t) \) is to be interpreted as the temperature field. After a sufficiently long time the temperature will always settle down to uniformly zero everywhere in the square. We will be concerned with transient solutions evolving from specified initial conditions on the field in the interior of the square.
First we must find the solution to the partial differential equation subject to the boundary conditions. We make the ansatz that the field can be written as a product of three functions, each a function of only one variable:

\[ \Psi(x, y, t) = X(x)Y(y)T(t) \]  

Inserting these into the differential equation and dividing through by \( XYT \) we obtain:

\[ \frac{T'}{T} = D \frac{X''}{X} + D \frac{Y''}{Y} \]  

(9.7)

Following a standard technique we see that under our ansatz the left hand side is only a function of \( t \) while the RHS is only a function of \( x \) and \( y \). The only consistent way out is for the LHS to be a constant, say \(-\lambda D\) (the separation constant). The reason for making the separation constant proportional to \( D \) is simply for convenience. Then:

\[ T' = -\lambda DT \]  

(9.8)

or

\[ T(t) = e^{-\lambda Dt} \]  

(9.9)

where we have not included the arbitrary coefficient since it will be taken into account later.

In our system we now have

\[ -\lambda D - D \frac{X''}{X} = D \frac{Y''}{Y} \]  

(9.10)

Now the LHS is a function only of \( x \) and the RHS is a function only of \( y \). We proceed as before by introducing a new separation constant:

\[ \frac{Y''}{Y} = -\alpha^2 \]  

(9.11)

where we have made the separation coefficient explicitly negative and absorbed the diffusion coefficient \( D \) for notational compactness. We can write immediately:

\[ Y''(y) = -\alpha^2 Y(y) \]  

(9.12)

which has solutions

\[ Y(y) = A\cos(\alpha y) + B\sin(\alpha y) \]  

(9.13)

We must use the boundary conditions if possible to decide about the coefficients \( A \) and \( B \). We see that the boundary conditions can be satisfied if

\[ \alpha = n\pi; \quad n = 1, 2, 3, \ldots \]  

(9.14)
9.1. DIFFUSION EQUATION

and if
\[ A = 0 \] (9.15)

Now we return to the \( X(x) \) equation
\[ \frac{X''}{X} = -\lambda + \alpha^2 \equiv -\gamma^2 \] (9.16)

Proceeding as before we find
\[ X(x) = \sin(\gamma x); \quad \gamma = m\pi; \quad m = 1, 2, 3, \ldots \] (9.17)

We can now see that
\[ \lambda = \alpha^2 + \gamma^2 \] (9.18)
\[ = (n^2 + m^2)\pi, \quad m, n = 1, 2, 3, \ldots \] (9.19)

We are at last in a position to write a solution to the system:
\[ \Psi(x, y, t) = A_{n,m} \sin(n\pi x) \sin(m\pi y) e^{-\lambda_{n,m} Dt} \] (9.20)

Each mode is labeled by the index \( n, m \) and decays exponentially toward the asymptotic solution of zero everywhere in the square. The higher index modes (smaller spatial scales) decay much more rapidly. We next turn to the question of how the mode amplitudes \( A_{n,m} \) are determined.

Suppose for an individual realization, the initial condition is given by a shape \( \Psi(x, y, 0) \). This initial shape should satisfy the boundary conditions. We can expand this function into a double Fourier series:
\[ \Psi(x, y, 0) = \sum_{n,m} B_{n,m} \sin(n\pi x) \sin(m\pi y) \] (9.21)

where the Fourier coefficients are obtained by multiplying through by \( \sin(n'\pi x) \sin(m'\pi y) \) and integrating from zero to one. We obtain
\[ \int \int \Psi(x, y, 0) \sin(n'\pi x) \sin(m'\pi y) \, dx \, dy = \frac{1}{4} B_{n',m'} \] (9.22)

Hence, the \( B_{n,m} \) can be obtained by a knowledge of the initial field.

9.1.1 Random Initial Conditions

Now suppose the \( B_{n,m} \) are random variables such that
\[ \langle B_{n,m} \rangle = 0 \] (9.23)
\[ \langle B_{n,m} B_{n',m'} \rangle = \sigma_{n,m}^2 \delta_{n,n'} \delta_{m,m'} \] (9.24)
The quantity $\sigma^2_{n,m}$ is a wavenumber spectrum of the initial conditions.

Now we look at realizations of the field after a time $t_0$. What is the wavenumber spectrum of the field at this time? We can find it by simply reducing the initial condition coefficients by the factor $e^{-\lambda_{nm} Dt_0}$. The spectrum is correspondingly reduced at each wavenumber by the factor $e^{-2\lambda_{nm} Dt_0}$.

The EOFs of the field at time $t_0$ are simply the functions

$$\psi_{n,m}(x,y) = 2 \sin(n\pi x) \sin(m\pi y) \quad (9.25)$$

They have the property:

$$\int \int \psi_{n,m}(x,y)\psi_{n',m'}(x,y) \, dx \, dy = \delta_{nn'}\delta_{mm'} \quad (9.26)$$

An objection to the above analysis is that it is not the way we usually examine a field. Usually we look at the evolution of a field over time instead of looking at a fixed time after a random initial condition.

### 9.2 Random Forcing of Waves

For variety we consider a wave equation for a string clamped at each end ($x = 0, 1$). The governing equation is

$$\frac{1}{c^2} \frac{\partial^2 \Phi}{\partial t^2} = \frac{\partial^2 \Phi}{\partial x^2} + F(x,t) \quad (9.27)$$

where $c$ is the velocity of waves and $F(x,t)$ is an external forcing function (the mad hammerer). From the outset let us define the forcing function as

$$F(x,t) = \sum_n A_n(t) \sqrt{2} \sin(n\pi x) \quad (9.28)$$

where the $A_n(t)$ are a set of stationary noise functions uncorrelated with one another and having spectral densities $S_n(f)$. Note that

$$\int_0^1 \sqrt{2} \sin(n\pi x) \cdot \sqrt{2} \sin(m\pi x) \, dx = \delta_{nm} \quad (9.29)$$

so that

$$A_n(t) = \int_0^1 \sqrt{2} \sin(n\pi x) F(x,t) \, dx \quad (9.30)$$
9.2. RANDOM FORCING OF WAVES

We proceed by multiplying through the governing equation by $e^{2\pi ift}$ and integrating over time:

$$-\left(\frac{2\pi f}{c}\right)^2 \Phi(x,f) = \frac{\partial^2 \tilde{\Psi}(x,f)}{\partial x^2} + \sum_n \tilde{A}_n(f) \sqrt{2} \sin(n \pi x)$$  \hspace{1cm} (9.31)

where the tilde has been used to denote the Fourier Transform in time. We expand $\tilde{\Psi}(x,f)$ into the basis set $\sqrt{2} \sin(m \pi x)$:

$$\tilde{\Psi}(x,f) = \sum_m \tilde{C}_m(f) \sqrt{2} \sin(m \pi x)$$  \hspace{1cm} (9.32)

where the coefficients are given by

$$\tilde{C}_m(f) = \int_0^1 \tilde{\Psi}(x,f) \sqrt{2} \sin(m \pi x) \, dx$$  \hspace{1cm} (9.33)

We can insert this representation in the governing equation. After integrating by parts twice on the RHS first term, we can multiply through by $\sqrt{2} \sin(m \pi x)$ and integrate from 0 to 1. This leaves

$$-\left(\frac{2\pi f}{c}\right)^2 \tilde{C}_m(f) = -(\pi m)^2 \tilde{C}_m(f) + \tilde{A}_m(f)$$  \hspace{1cm} (9.34)

After rearranging:

$$\tilde{C}_m(f) = \frac{\tilde{A}_m(f)}{(\pi m)^2 - \left(\frac{2\pi f}{c}\right)^2}$$  \hspace{1cm} (9.35)

Now we know. The (frequency dependent) EOFs for the string being agitated by the mad hammer are

$$EOF_n(f) = \sqrt{2} \sin(m \pi x) e^{2\pi ift}$$  \hspace{1cm} (9.36)

The eigenvalue spectrum for these is

$$\lambda_n(f) = \frac{S_n(f)}{\left((\pi m)^2 - \left(\frac{2\pi f}{c}\right)^2\right)^2}$$  \hspace{1cm} (9.37)

Note that the eigenvalue spectrum has a singularity along the straight line

$$f = cm/2$$  \hspace{1cm} (9.38)

in the $(f-m)$ plane. The slope is proportional to the wave speed. The singularity is because each mode $m$ has a ‘natural’ frequency $cm/2$. A small amount of damping $(\partial \Phi/\partial t)$ would make the singularities finite.
Chapter 10

Combining Data Optimally

In climatology we often encounter the estimation problem. It comes in many different forms. We basically are trying to infer the value of some quantity from an incomplete record of evidence. For example we have a time series running over the past and up to the present. We might like to predict what will happen after the next time step (prediction or extrapolation). Another example is that there might be a missing element in the time series and we would like to fill it in in some optimal sense and have an idea about the confidence interval associated (interpolation). We might have two estimates of a parameter and wish to know how to combine the estimates in such a way as to improve on the estimates of either taken alone (optimal combination). We might wish to know how to filter a data stream in which there is a mixture of natural variability and signal in such a way as to remove as much as possible the natural variability (optimal filtering). There are many others.

10.1 Combining Two Estimates

Suppose we are trying to estimate the temperature of a body based upon two noisy types of data. We might insert a (poor) thermometer as one estimate ($\hat{T}_1$) and look at the body with an infrared radiometer ($\hat{T}_2$) as another presumably independent estimate. Let $T$ be the true temperature and let us assume that each measurement is unbiased:

$$T = \langle \hat{T}_1 \rangle = \langle \hat{T}_2 \rangle$$

(10.1)
Consider next an estimate consisting of a linear combination of the two independent estimates:

\[ \hat{T} = \alpha_1 \hat{T}_1 + \alpha_2 \hat{T}_2 \] 

(10.2)

To assure that the new estimate is still unbiased we must require

\[ \alpha_1 + \alpha_2 = 1 \] 

(10.3)

Now we may form the mean square error (MSE):

\[ \epsilon^2(\alpha_1, \alpha_2) = \langle (T - \hat{T})^2 \rangle \] 

(10.4)

\[ = \langle (T - \alpha_1 \hat{T}_1 - \alpha_2 \hat{T}_2)^2 \rangle \] 

(10.5)

\[ = \langle (\alpha_1(T - \hat{T}_1) + \alpha_2(T - \hat{T}_2))^2 \rangle \] 

(10.6)

\[ = \alpha_1^2 \langle (T - \hat{T}_1)^2 \rangle + \alpha_2^2 \langle (T - \hat{T}_2)^2 \rangle \] 

(10.7)

\[ = \alpha_1^2 \sigma_1^2 + (1 - \alpha_1)^2 \sigma_2^2 \] 

(10.8)

where \( \sigma_i^2 = \text{Var}(\hat{T}_i) \) and we have assumed the covariance \( \text{Cov}(\hat{T}_1, \hat{T}_2) = 0 \).

Now if we take the derivative of \( \epsilon^2 \) with respect to \( \alpha_1 \) and set it to zero we get

\[ \alpha_i = \frac{1}{\sigma_i^2}, \quad i = 1, 2 \] 

(10.9)

Hence, each observation is to be weighted inversely by the variance of that particular measurement. Then measurements that are poor (large variance) will be weighted less than those that are more certain (small variance).
Chapter 11

Optimal Interpolation

The technique was introduced into meteorology by Gandin (1963). In short range forecasting it is necessary to initialize numerical models. Data are gathered at irregular points in the space, but models run on regular grids. This means that in the feeding of initial data into the forecast cycle, one must first interpolate the data from the irregular station grid ($r_s, s = 1 \ldots$) to the regular model grid ($r_m, m = 1 \ldots$). We illustrate the technique of optimal interpolation by assuming that we know the statistics of the field in question, say $T(r)$, from climatology. In practice, these background statistics might come from a previous forecast rather than climatology or perhaps even a blend of the two.

We seek an estimate of $T(r_m)$, which we call $\hat{T}_m$ which is a linear combination of the data taken from the stations:

$$\hat{T}_m = T_m^0 + \sum_{s=1}^{s} w_{ms} (T_s - T_s^0)$$  \hspace{1cm} (11.1)

where $T_m^0$ is the climatological value of $T$ at $r_m$, $T_s^0$ is the climatological value of $T$ at $r_s$, and $T_s$ is the observed value of $T$ at $r_s$. The coefficients $w_{ms}$ are optimal weights associated with the interpolation. They are to be chosen such that the mean square error (MSE) is least. We form the MSE by

$$\epsilon^2 = \langle (\hat{T}_m - T_m)^2 \rangle$$  \hspace{1cm} (11.2)

$$= \langle (T_m^0 + \sum_{s=1}^{s} w_{ms} (T_s - T_s^0) - T_m)^2 \rangle$$  \hspace{1cm} (11.3)

$$= \langle (T_m^0 - T_m)^2 \rangle + 2 \sum_{s} w_{ms} \langle (T_s - T_s^0)(T_m^0 - T_m) \rangle$$  \hspace{1cm} (11.4)

133
\[
+ \sum_s \sum_{s'} w_{ms} w_{ms'} \langle (T_s - T_s^0)(T_{s'} - T_{s'}^0) \rangle \quad (11.5)
\]

The least MSE results when

\[
\frac{\partial \epsilon^2}{\partial w_{ms}} = 0 \quad (11.6)
\]

Which results in the set of linear equations:

\[
\sum_s C_{s's} w_{ms} = -G_{s'm} \quad (11.7)
\]

where

\[
C_{s's} = \langle (T_{s'} - T_{s'}^0)(T_s - T_s^0) \rangle \quad (11.8)
\]

\[
G_{s'm} = \langle (T_{s'} - T_{s'}^0)(T_m - T_m^0) \rangle \quad (11.9)
\]

Once the \( C_{s's} \) and \( G_{s'm} \) are known the linear equations for the optimal weights can be solved. This matrix inversion must be conducted for each grid point. But once all the inversions are performed the interpolation weights \( w_{ms} \) can be tabulated and stored. Then for a set of observations at the stations we can find the interpolated values at the model grid points by use of (11.1).

To simplify matters one often uses a model form of the covariances of the field between distant points. One of the simplest is due to Thiebaux (1976):

\[
\rho(r) = \left[ \cos cr + \frac{\sin cr}{Lc} \right] e^{-r/L} \quad (11.10)
\]

where \( c \) and \( L \) are empirical coefficients which will depend on location and season. This particular model is isotropic, but in fact the correlation structure in midlatitudes is not isotropic.

References


Chapter 12

Optimal Detection

12.1 Introduction

There are a number of optimal approaches being considered in the problem of detecting forced climate change (e.g., Hasselmann, 1979; Bell, 1982, 1986; Hasselman, 1993; North et al., 1995, North and Kim, 1995; Sauter et al., 1996; Hegerl et al., 1996; Stevens and North, 1996). For the most part the mathematical framework is identical (Hegerl and North, 1996), but the ways the techniques are being applied and the data being invoked differ widely depending upon the research group. It is important that a variety of approaches be used since if they arrive at similar conclusions it adds to the force of the argument that signals are really being detected in the various data streams. In fact, the acceptance of such a conjecture (if it is true) will come only gradually as a diversity of evidence accumulates.

Part of the problem of acceptance of the optimal detection theory framework is its mathematical complexity. For an assertion about anthropogenically induced climate change to stick, it must be understandable by the average scientist in neighboring disciplines. The works of Norbert Wiener (e.g., Wiener, 1945?) and later treatises on space-time signal processing are neither familiar nor are they easy reading for the average physical scientist. One of our goals in this paper is to rederive the optimal detection procedure in a simple framework that will make it accessible and intuitive. As so often happens after a laborious, rigorous derivation, one finds a simple geometrical picture of the process that sheds new light on it and even provides a new and better interpretation. This is the case and we spell it out in Section 2.

An aspect of the studies that is only recently coming into play is the
inclusion of several signals in the detection problem. It appears that in the anthropogenic signal case, one must include both the greenhouse gas forcing along with the anthropogenic aerosol forcing. There are at least two other significant deterministic signals, the volcanic and the solar variability. We call them deterministic since we know the record of the forcing and as we look at an ensemble of simulated records we can include these as though they were deterministic. We approach the problem of multiple signals by considering the length of the component of the signal which is perpendicular to the sum of the interfering signals. We also look at the theoretical signal-to-noise ratio (TSNR), which is the a priori assessment based only upon model-dependent characteristics of the procedure, for the composite signal with each individual added one at a time.

We proceed to use a simple climate model to provide signal waveforms whose strengths are to be determined by the data. The optimal filter (linear procedure whereby the data are reduced) is constructed using a variety of state of the art climate models. Then we use the 100 yr stream of surface temperature data at 36 locations on the Earth and apply the filter to it to estimate the various signal strengths along with assessments of the performance of the filter in each case.

12.2 \( K \) independent estimators

Consider estimating the temperature of a reservoir with two devices. Let the estimators \( \hat{T}_1 \) and \( \hat{T}_2 \) be unbiased; i.e., \( \langle \hat{T}_1 \rangle = \langle \hat{T}_2 \rangle = T \) where \( T \) is the true temperature, and \( \langle \cdot \rangle \) means ensemble average.

The individual estimators are assumed to be of the form \( \hat{T}_i = T + \epsilon_i \); where the errors \( \epsilon_i, i = 1, 2 \) are assumed to be random variables taking on different values in each realization of the measurement process. The errors or noise are assumed to have mean zero when considered over a large number of trials: \( \langle \epsilon_i \rangle = 0 \), and the covariances of the errors are given by: \( \langle \epsilon_i \epsilon_j \rangle = \sigma_i^2 \delta_{ij}; i, j = 1, 2 \). The previous expression states that the errors of the separate devices are assumed to be uncorrelated and that their individual variances are given by \( \sigma_1^2 \) and \( \sigma_2^2 \). We assume that these characteristics of the errors are known beforehand. Our task is to take one realization of the measurement process and obtain an optimal estimate of the true reservoir temperature. We wish to make maximal use of the data collected from each device in an appropriate linear combination. The question is, what is the appropriate weighting to assign to each measurement? We form the estimate:

\[
\hat{T} = W\hat{T}_1 + (1 - W)\hat{T}_2
\]  

(12.1)
where $W$ is a weight to be adjusted to make the MSE least. The estimator $\hat{T}$ is clearly unbiased if the individual estimators are. We can form the mean square error (MSE) for the measurement as

\[
\epsilon^2 = \langle (\hat{T} - T)^2 \rangle = W^2(\sigma_1^2 + \sigma_2^2) - 2W\sigma_2^2 + \sigma_2^2
\]

(12.2)

(12.3)

The latter is a quadratic in $W$ and is shown in Fig. 1 for a choice of $\sigma_1^2 = 2\sigma_2^2 = 1$. The point of this figure is that the MSE is rather insensitive to the choice of $W$ so long as it is near its optimum value. This is an important point to be stressed later in the climate signal detection exercises.

The minimum of the quadratic above is easily found, and it yields the familiar and very important result:

\[
\hat{T}_{opt} = \frac{1}{\eta^2} \left( \frac{\hat{T}_1}{\sigma_1^2} + \frac{\hat{T}_2}{\sigma_2^2} \right)
\]

(12.4)

where

\[
\eta^2 = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}
\]

(12.5)

The result is easily generalized to include $K$ independent unbiased estimators:

\[
\hat{T}_{opt} = \frac{1}{\eta^2} \sum_{k=1}^{K} \frac{\hat{T}_k}{\sigma_k^2}
\]

(12.6)

and

\[
\eta^2 = \sum_{k=1}^{K} \frac{1}{\sigma_k^2}
\]

(12.7)

After some algebra it can be shown that the optimal MSE is just

\[
\epsilon_{opt}^2 = \frac{1}{\eta^2} = \frac{1}{\sum_{k=1}^{K} \frac{1}{\sigma_k^2}}
\]

(12.8)

This shows that adding another device always lowers the MSE no matter how poor its quality.

The derivation presented above required that the individual errors be uncorrelated with one another. If this were not so the coordinate axes could simply be rotated to the principle axes of the error or noise covariance matrix. Then the entire formalism goes through as before except in the rotated coordinate system. In climatology this is the tranformation to the Empirical Orthogonal Function (EOF) basis set.
In two dimensions this is easily spelled out explicitly. Let the covariance matrix of the noise be given by
\[ C_{ij} = \langle \epsilon_i \epsilon_j \rangle \]  
(12.9)

Taking the noise to be distributed bivariate normally, the contours of equal probability of occurrence of pairs of values of \((\epsilon_1, \epsilon_2)\) are given by (p126, Thiebaux, 1994):
\[ \frac{T_1^2}{\sigma_1^2} - 2\rho \frac{T_1 T_2}{\sigma_1 \sigma_2} + \frac{T_2^2}{\sigma_2^2} = \text{constant} \]  
(12.10)

which is an ellipse in the \((T_1, T_2)\) plane. In this two dimensional case we can find an angle \(\theta\) to rotate the coordinate axes to, such that the principal axes of the ellipse coincide with new coordinate axes \((T'_1, T'_2)\). In the new coordinate system, \(T'_1\) and \(T'_2\) are uncorrelated. In the case of \(K\) dimensions, the figure is an ellipsoid in the \(K\) dimensional space and a simple length preserving rotation can also be used to find the appropriate coordinate system.

The simple derivation of optimal weighting of independent estimators is familiar to many atmospheric scientists. The result is very intuitive. We simply weight each estimator inversely according to its individual error variance.

### 12.2.1 Characterizing a signal

Consider a sinusoidal wave in one dimension
\[ s(t) = s^{(+) \frac{1}{\sqrt{2}}} \cos \omega t + s^{(-) \frac{1}{\sqrt{2}}} \sin \omega t \]  
(12.11)

The coefficients \(s_1\) and \(s_2\) determine the amplitude and phase of the wave. We can characterize the signal as a vector in a two dimensional space:
\[ \mathbf{S} = s_1 \mathbf{e}_1 + s_2 \mathbf{e}_2 \]  
(12.12)

where \(\mathbf{e}_i\) is a unit vector in the \((1,2)\) plane along the \(i\)th axis; \(i = 1, 2\). The above result can of course be generalized to any number of dimensions, if for example, the signal is composed of many harmonics. For example,
\[ S(t) = s_0 + \sum_{n=1}^{n=N} \left( s^{(+) \frac{1}{\sqrt{2}}} \cos n\omega_0 t + s^{(-) \frac{1}{\sqrt{2}}} \sin n\omega_0 t \right) \]  
(12.13)

Then the vector representation of \(S(t)\) is
\[ \mathbf{S} = (s_0; s_1^{(+) \frac{1}{\sqrt{2}}}, s_1^{(-) \frac{1}{\sqrt{2}}}; s_2^{(+) \frac{1}{\sqrt{2}}}, s_2^{(-) \frac{1}{\sqrt{2}}} \ldots) \]  
(12.14)
or
\[ \mathbf{S} = \sum_{n=0}^{N} \left( s_n^{(+)} \mathbf{e}_n^{(+)} + s_n^{(-)} \mathbf{e}_n^{(-)} \right) \]  
(12.15)

Hence, if the signal is composed of \( N \) harmonics, there will be \( 2N + 1 \) coefficients representing the amplitude and phase of each, with the exception of the zero frequency harmonic which has no phase.

### 12.2.2 Signal and noise

The additive noise can also be decomposed into frequency components (dropping the awkward (+) and (-) superscripts)
\[ \mathbf{N} = n_1 \mathbf{e}_1 + n_2 \mathbf{e}_2 \]  
(12.16)

In this case the components \( n_1, n_2 \) are random variables and uncorrelated. (If they were correlated, we would rotate the axes to new coordinates such that there is no correlation; i.e., in climate we would use the EOF basis set.) For each realization of the process a new value of \( n_1 \) and \( n_2 \) must be drawn from a distribution function and the distribution of \( n_1 \) and that of \( n_2 \) are independent. If the same frequency component of noise is added to that of the signal, we can write
\[ \mathbf{D} = (s_1 + n_1) \mathbf{e}_1 + (s_2 + n_2) \mathbf{e}_2 \]  
(12.17)

where we have used \( \mathbf{D} \) to indicate “data”. It is worth noting that if the noise process is a stationary time series, the noise from one frequency component to another is uncorrelated. Hence, in this simple case no rotation is required.

In all the applications that follow we assume the signals are linearly added to one another and to the natural variability background (the “noise”). We are now in a position to form some estimators of interesting quantities. For example, an unbiased estimator of \( s_1 \) is simply \( \mathbf{e}_1 \cdot \mathbf{D} \), since \( \langle n_1 \rangle = 0 \) and thus \( \langle \mathbf{e}_1 \cdot \mathbf{D} \rangle = s_1 \).

### 12.2.3 Fingerprint estimator of \( \mathbf{S} \)

A common problem in climatology is that we know the waveform of the signal (in the above example the frequency and phase) but want to know its strength. In other words we know the direction of the signal vector (indicated by the unit vector \( \mathbf{e}_S \equiv \mathbf{S}/|\mathbf{S}| \)). An unbiased estimator of \( \mathbf{S} \equiv |\mathbf{S}| \) is
\[ \hat{S}_{rf} = \mathbf{e}_S \cdot \mathbf{D} \]  
(12.18)
where the subscript \( rf \) indicates ‘raw fingerprint’. In other words, we find the length of the component of the data vector which lies along the direction of \( S \). The raw fingerprint estimator has a MSE

\[
\epsilon_{r f}^2 = \frac{s_1^2 \sigma_1^2 + s_2^2 \sigma_2^2}{s_1^2 + s_2^2} \quad (12.21)
\]

The raw fingerprint method is very easy to implement and has attracted some users. On the other hand, it does not take advantage of the fact that the \( \sigma_1^2 \) and \( \sigma_2^2 \) might be quite different. Hence, we might want to weigh the information from the two component directions optimally. The way to do this is presented in the next subsection.

### 12.2.4 Optimal weighting

Consider a two dimensional case in which we do know the direction of the signal vector \((\mathbf{e}_S)\) and it makes an angle \( \theta_1 \) with the \( \mathbf{e}_1 \) axis. Then we can write for the component of \( S \) along the \( \mathbf{e}_1 \) axis:

\[
s_1 = |S| \cos \theta_1 = S \cos \theta_1 = S \mathbf{e}_1 \cdot \mathbf{e}_S \quad (12.22)
\]

or

\[
S = \frac{s_1}{\mathbf{e}_1 \cdot \mathbf{e}_S} \quad (12.23)
\]

If there were no noise we could calculate \( S \) by first obtaining its component in the 1 direction from data, then dividing by the direction cosine of the known signal vector and the 1 axis. This means we can form an unbiased estimate of \( S \): \( \hat{S}^{(1)} = \frac{\mathbf{e}_1 \cdot \mathbf{D}}{\mathbf{e}_1 \cdot \mathbf{e}_S} \) (note that \( \langle \hat{S}^{(1)} \rangle = S \)). Hence, the data vector is to be projected along the 1-axis and inversely weighted by the direction cosine of the signal vector to the 1-axis. This unbiased estimator of \( S \) has an error variance of

\[
\epsilon_1^2 = \frac{\sigma_1^2}{(\mathbf{e}_1 \cdot \mathbf{e}_S)^2} \quad (12.27)
\]
But we have many statistically independent unbiased estimators of $S$, one for each component direction. The problem has been reduced to the same one as the thermometers in the reservoir analysed at the beginning of this section. Hence, the optimal estimator of $S$ is

$$\hat{S}_{opt} = \frac{1}{\eta^2} \sum_{k=1}^{K} \frac{(e_S \cdot e_k)(e_k \cdot D)}{\sigma_k^2}$$

(12.28)

$$= \left\{ \frac{1}{\eta^2} \sum_{k=1}^{K} \frac{(e_S \cdot e_k)e_k}{\sigma_k^2} \right\} \cdot D$$

(12.29)

with

$$\eta^2 = \sum_{k=1}^{K} \frac{(e_S \cdot e_k)^2}{\sigma_k^2}$$

(12.30)

In the last expression for $\hat{S}_{opt}$ we show the data vector $D$ factored out to emphasize that the procedure is a linear operation or projection of the data vector; hence, the term filter. Except for a slight change in notation this is the same formula derived in North et al. (1995) and used in North and Kim (1995) and Stevens and North (1996). In the earlier papers $\sigma_i^2 \equiv \lambda_i$, since this variance is the eigenvalue of the corresponding EOF; we also used $\eta = \gamma/|S|$. In the earlier paper we interpreted $\gamma$ as the signal-to-noise ratio, whereas $\eta$ can be thought of as the inverse of the standard error (unit signal magnitude).

Let us recall the assumptions. First and foremost, we assumed the linear additivity of the signal and the noise. We have used the principle component directions to formulate the problem from the beginning. That is, we chose the coordinate axes to be the principle axes of the covariance ellipsoid of the noise vector. We had to assume knowledge of the direction of the signal waveform. Our job is to estimate its strength given this information. The quantity $\eta^2$ is a good a priori measure of the quality of the procedure, since for a signal strength of unity, it is the signal-to-noise ratio squared. We can use $\eta^2$ as computed with models to tell which vector components are most important in the estimation problem without really invoking the data. This is very important since we can use our climate models to condition our choice of the subspace within which we can make a reliable estimation of signal strength without involving the data itself (cheating).

Consider the error involved in the use of imperfect models in constructing the filter. The first type of error is in establishing an incorrect fingerprint. In the present context this means the vector $e_S$ has the wrong direction in the state space. An equivalent statement is that the direction cosines $e_S \cdot e_i$ are incorrect. The single constraint is that the squares of
the direction cosines must add up to unity. An incorrect fingerprint can lead to a bias in the estimation of the signal strength. For this reason, it is well to find ways to eliminate aspects of the model predicted signal which may lead to incorrect signal waveform prediction. This could be done by eliminating certain subspaces of the state space, but this is probably not a good approach since the EOFs are very irregular functions over the globe and it is not easy to relate these shapes to the areas that we know are weak in the signal generation. Instead, it might be better to mask off certain regions on the globe, such as the polar regions where we know the models perform poorly. Once we have masked off certain areas (with tapered edges) we completely redo the problem including the EOFs on the newly masked planet. We do not pursue this in the present work but it is worthy of serious consideration in future studies.

Another type of error comes from the optimal weights as generated from models. This type of error is less egregious than error in the signal waveform. Since the estimator is composed of $K$ independent estimators which are assumed to be unbiased, the weighting does not introduce a bias. If erroneous they can lead to a suboptimal estimator and they can lead to an underestimation of the theoretical mean squared error ($1/\eta^2$). It turns out that since the minimum in the mean squared error as a function of the weights is the minimum of a multidimensional parabolic surface (actually the intersection of this parabolic surface with the plane $\sum_i w_i = 1$), the MSE is not sensitive to the exact choice of the weights ($\Delta r^2 \sim (\Delta W)^2$).

### 12.2.5 Raw fingerprint estimator

Having derived the optimal estimator for signal strength, we return to the raw fingerprint estimator in the same notation to make a straightforward comparison of the two filters. The estimator based upon the raw fingerprint projection can be written

$$\hat{S}_{rf} = (\mathbf{e}_1 \cdot \mathbf{e}_S)^2 \left( \frac{\mathbf{e}_1 \cdot \mathbf{D}}{\mathbf{e}_1 \cdot \mathbf{e}_S} \right) + (\mathbf{e}_2 \cdot \mathbf{e}_S)^2 \left( \frac{\mathbf{e}_2 \cdot \mathbf{D}}{\mathbf{e}_2 \cdot \mathbf{e}_S} \right)$$

$$\hat{S}_{rf} = \left\{ (\mathbf{e}_1 \cdot \mathbf{e}_S)^2 \left( \frac{\mathbf{e}_1}{\mathbf{e}_1 \cdot \mathbf{e}_S} \right) + (\mathbf{e}_2 \cdot \mathbf{e}_S)^2 \left( \frac{\mathbf{e}_2}{\mathbf{e}_2 \cdot \mathbf{e}_S} \right) \right\} \cdot \mathbf{D}$$

Or for $K$ dimensions:

$$\hat{S}_{rf} = \left\{ \sum_{k=1}^{K} (\mathbf{e}_k \cdot \mathbf{e}_S)^2 \left( \frac{\mathbf{e}_k}{\mathbf{e}_k \cdot \mathbf{e}_S} \right) \right\} \cdot \mathbf{D}$$

Note that the raw fingerprint filter makes no use of the variance (natural variability) information in its projection of the data.
12.3. INTERFERING SIGNALS

Table 12.1: Vectors denoting deterministic signals in the climate system.

<table>
<thead>
<tr>
<th>Signal</th>
<th>Vector notation</th>
<th>Unit Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greenhouse Gases</td>
<td>$G$</td>
<td>$e_G \equiv G/</td>
</tr>
<tr>
<td>Anthropogenic Aerosols</td>
<td>$A$</td>
<td>$e_A$</td>
</tr>
<tr>
<td>Volcanic Aerosols</td>
<td>$V$</td>
<td>$e_V$</td>
</tr>
<tr>
<td>Solar Cycle</td>
<td>$S$</td>
<td>$e_S$</td>
</tr>
</tbody>
</table>

12.3 Interfering Signals

In the climate problem there are many deterministic climate signals present in the data stream along with the natural variability (climate noise). For example, there are the greenhouse warming signal $G$, the volcanic signal (stratospheric aerosols) $V$, the anthropogenic tropospheric aerosol signal $A$ and the solar cycle signal $S$. If we want to estimate the strength (length) of $S$ in the presence of another of these signals whose strength is not known exactly, we will encounter problems if the vectors are nearly parallel or anti-parallel.

12.3.1 One interfering signal

If the direction of the desired signal (say $S$) and interfering signal (say $V$) are known, we can obtain independent estimates of their strengths by estimating the components of each which are perpendicular to the other. For example, consider the component of $S$ which is perpendicular to $V$:

\[
S_{\perp V} = (1 - e_V e_V) \cdot S
\]

\[
= S - e_V (e_V \cdot S)
\]

where $e_V \equiv V/|V|$ is a unit vector along $V$. Hence, using this projection procedure (operator) we can now proceed to estimate the strength of $S_{\perp V}$ and therefore find the strength of $S$, since $S = |S_{\perp V}|/\sqrt{1 - (e_S \cdot e_V)^2}$. The problem of course is that $S_{\perp V}$ will be shorter than $S$ with a corresponding loss of performance (signal-to-noise ratio $= \eta S \cdot e_V$) in the procedure.

We can now use the same procedure to find the length of $V_{\perp S}$ and therefore the length of $V$. As a consistency check we could then proceed to look at the sum of the parallel components, which in principle are now known.

\[
S_{\parallel V} = e_V \cdot S
\]
\( \mathbf{V}_{\parallel S} = \mathbf{e}_S \cdot \mathbf{V} \)  

(12.37)

It is obviously of interest to know the angle between \( \mathbf{S} \) and \( \mathbf{V} \),

\[ \theta_{S,V} = \arccos(\mathbf{S} \cdot \mathbf{V}) \]  

(12.38)

If the length and direction of the interfering signal are both known, we have an unbiased estimator of the length of \( \mathbf{S}_{\perp V} \), which when divided by \( \sqrt{1 - (\mathbf{e}_S \cdot \mathbf{e}_V)^2} \) becomes an unbiased estimator of \( S \). We can optimally combine this with the independent estimate based upon the parallel component which can be found by first subtracting (the known) \( \mathbf{V}_{\parallel S} \) from the data stream.

### 12.4 Examples from 100 years of Temperature Data

Stevens and North (1996) have analyzed 100 years of surface temperature data based upon 36 \( 10^\circ \times 10^\circ \) boxes distributed over the earth as shown in Fig.XX. The data come from the UK data set of month averages and the boxes were chosen for their continuity, representativeness and distribution. In each box there are at least 4 continuous time series. In our time series we excluded the winter half year in the extratropics and kept the entire annual average in the tropics as explained in Stevens and North (1996). In future work we intend to keep all months using a cyclostationary EOF basis set (Kim et al., 1996). The retained data form a 36 component multivariate discrete time series. In the absence of external forcing the time series should be stationary as we have found in 1000yr time series (control runs) with several coupled ocean-atmosphere climate models.

By analyzing these long control runs we were able to find the spatial EOFs for each Fourier frequency component (actually, we took the spatial factors to be the same over the narrow frequency band employed in the previous study as well as here). Since our original interest was in detecting the solar signal amidst climate noise and other deterministic signals we restrict our investigation to the “solar band”, which includes only the 8 frequencies: \( 1/6, 1/7, 1/8, 1/9, 1/10, 1/11, 1/12, 1/13 \text{yr}^{-1} \). The corresponding time scales range from 16.67 yr to 7.69 yr. The band chosen here is purposely designed to exclude as much ENSO activity as possible (\( f > 1/5 \text{yr}^{-1} \)) because the climate models being used do not simulate ENSO very faithfully. The low frequency cutoff was to exclude interference with the long term trends in climate as much as feasible. An additional advantage offered by the low frequency cutoff is that the temporal eigenfunctions for the 100 yr
Table 12.2: Angles in degrees between pairs of the four signal vectors in the 546 dimensional space.

<table>
<thead>
<tr>
<th>vector pair</th>
<th>angle (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathbf{S} \cdot \mathbf{V} )</td>
<td>88.0</td>
</tr>
<tr>
<td>( \mathbf{S} \cdot \mathbf{G} )</td>
<td>77.9</td>
</tr>
<tr>
<td>( \mathbf{S} \cdot \mathbf{A} )</td>
<td>101.0</td>
</tr>
<tr>
<td>( \mathbf{V} \cdot \mathbf{G} )</td>
<td>93.8</td>
</tr>
<tr>
<td>( \mathbf{V} \cdot \mathbf{A} )</td>
<td>84.2</td>
</tr>
<tr>
<td>( \mathbf{G} \cdot \mathbf{A} )</td>
<td>153.3</td>
</tr>
<tr>
<td>( \mathbf{S} \cdot (\mathbf{V} + \mathbf{G} + \mathbf{A}) )</td>
<td>80.9</td>
</tr>
<tr>
<td>( \mathbf{V} \cdot (\mathbf{S} + \mathbf{G} + \mathbf{A}) )</td>
<td>90.9</td>
</tr>
<tr>
<td>( \mathbf{G} \cdot (\mathbf{V} + \mathbf{S} + \mathbf{A}) )</td>
<td>116.9</td>
</tr>
<tr>
<td>( \mathbf{A} \cdot (\mathbf{V} + \mathbf{G} + \mathbf{S}) )</td>
<td>133.6</td>
</tr>
</tbody>
</table>

segment are closer to sines and cosines if the frequency is high compared to the \((\text{segment length})^{-1}\) (see the appendix of North and Kim, 1995). In the previous study (Stevens and North, 1996) we retained the \(1/5\) yr\(^{-1}\) harmonic. Each frequency has a sine and cosine component leading to \(8 \times 2 = 16\). Combining with the 36 stations gives us a climate system subspace of dimension 546. The vectors \(\mathbf{S}, \mathbf{N}, \mathbf{V}, \text{etc.}\), each have 546 components in this study.

A final important argument for using only the narrow band of frequencies in our study is that present coupled ocean/atmosphere climate models may not be able to produce the low frequency spectrum of fluctuations with any reliability. In the band of interest here even a mixed layer model appears to be adequate. Fluctuations in this band from the fully coupled GFDL model appear to be very similar to those in the corresponding mixed layer model.

We compute the signals \(\mathbf{S}, \mathbf{V}, \mathbf{G}, \text{and} \mathbf{A}\) from an energy balance model described by Stevens and North (1996) and elaborated on in more detail by Stevens (1996). Stevens (1996) also describes how the individual forcings are included. The statistics of the stochastic noise vector \(\mathbf{N}\) are not needed in computation of the signals, but are only needed later in the optimal weighting. Table 1 shows the angles between the various signal vectors in the 546 dimensional (global) space.
12.5 Results

Our main goal is to find best estimates of the lengths of $S, V, A$ and $G$. What we really find is the ratio of the length in the data to that in the EBM used to generate the input signal wave form. A plot of the global average of each of these versus time is shown in the Fig. 2. In the following subsections we present the results found by forming the perpendicular component of the individual signal to the sum of the other three. In this way we form an estimate of each individual signal’s strength.

In the tables which follow we show the results from using 4 climate models: EBM: optimal weights based upon a 10,000 yr run of our noise forced energy balance model; G/ML: optimal weights based upon a 1000 yr run of the GFDL atmospheric model with a mixed layer ocean; G/C: optimal weights based upon a 1000 yr run of the GFDL coupled atmosphere/ocean model; MPI: optimal weights based upon 1000 yr run of the Max Planck Institute coupled ocean/atmosphere model. In the tables GLOBAL means use of all 36 stations in the map. TROPICS means we used only the 18 stations in the tropics. N HEM means we used only the 18 stations in the Eastern Hemisphere and similarly for W HEM, S HEM and N HEM (see Stevens and North, 1996). In each of these cases we recomputed the optimal weights from scratch.

Looking through the tables one finds remarkable agreement between the 4 models used in computing optimal weights. The global case should be the most accurate since it takes into account more stations. However, there is usually pretty good agreement scanning the vertical columns for both TSNR and the amplitudes ($S, V, G$ and $A$). The column headed TSNR is the value of $\eta \times |S_{ebm}|$ for the particular set of stations and the model used in computing the optimal weights. The value of TSNR reflects the theoretical signal to noise ratio in the estimation assuming a signal strength given by the response of the EBM to the given forcing and its hypothesized strength.

12.5.1 Volcanic signal

The results for the volcanic signal are highly significant with a TSNR of over 3.0 with all models and for most subsets of the data. This means the signal strength is non zero with a probability of greater than 99%. Its strength averages 1.12 for all the global data across the 4 models. We also computed $\alpha$ for the volcanic signal individually perpendicular to the different signals (details not shown in this paper). The TSNR was similarly large and the value of $\alpha$ was again just above unity.
12.5. RESULTS

Table 12.3: Angles (deg) between various signal vectors in the 546 dimensional state space.

<table>
<thead>
<tr>
<th>vector pair</th>
<th>angle (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(S \cdot V)</td>
<td>88.0°</td>
</tr>
<tr>
<td>(S \cdot G)</td>
<td>77.9</td>
</tr>
<tr>
<td>(S \cdot A)</td>
<td>101.0</td>
</tr>
<tr>
<td>(V \cdot G)</td>
<td>93.9</td>
</tr>
<tr>
<td>(V \cdot A)</td>
<td>84.2</td>
</tr>
<tr>
<td>(G \cdot A)</td>
<td>153.3</td>
</tr>
<tr>
<td>(S \cdot (V + G + A))</td>
<td>80.9</td>
</tr>
<tr>
<td>(V \cdot (S + G + A))</td>
<td>90.9</td>
</tr>
<tr>
<td>(G \cdot (V + S + A))</td>
<td>116.9</td>
</tr>
<tr>
<td>(A \cdot (V + G + S))</td>
<td>133.6</td>
</tr>
</tbody>
</table>

Table 12.4: Volcanic signal estimated from its perpendicular component to the sum \(G + A + S\). TSNR is a theoretical signal to noise ratio for a signal whose strength equals that of the EBM generated signal. The statistic \(\alpha\) is the ratio of the estimated strength to that of the EBM generated signal. The SNR based upon the data is TSNR \(\times \alpha\).

<table>
<thead>
<tr>
<th>EBM</th>
<th>G/ML</th>
<th>G/C</th>
<th>MPI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TSNR (\alpha)</td>
<td>TSNR (\alpha)</td>
<td>TSNR (\alpha)</td>
</tr>
<tr>
<td>GLOBAL</td>
<td>4.19 1.33</td>
<td>4.10 0.93</td>
<td>4.24 1.20</td>
</tr>
<tr>
<td>TROPICS</td>
<td>4.03 1.35</td>
<td>4.32 1.04</td>
<td>4.14 1.12</td>
</tr>
<tr>
<td>EXTROPICS</td>
<td>3.24 1.37</td>
<td>2.74 1.30</td>
<td>2.83 1.06</td>
</tr>
<tr>
<td>N HEM</td>
<td>3.86 1.27</td>
<td>3.88 1.06</td>
<td>3.82 0.88</td>
</tr>
<tr>
<td>S HEM</td>
<td>2.95 1.31</td>
<td>3.18 1.16</td>
<td>3.19 1.11</td>
</tr>
<tr>
<td>W HEM</td>
<td>3.15 1.34</td>
<td>3.94 1.14</td>
<td>3.77 1.26</td>
</tr>
<tr>
<td>E HEM</td>
<td>2.81 1.23</td>
<td>3.46 1.12</td>
<td>3.65 1.06</td>
</tr>
</tbody>
</table>
Table 12.5: Solar signal estimated from its perpendicular component to the sum $G + A + V$. TSNR is a theoretical signal to noise ratio for a signal whose strength equals that of the EBM generated signal. The statistic $\alpha$ is the ratio of the estimated strength to that of the EBM generated signal. The SNR based upon the data is TSNR$\times\alpha$.

<table>
<thead>
<tr>
<th></th>
<th>EBM</th>
<th>G/ML</th>
<th>G/C</th>
<th>MPI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TSNR</td>
<td>$\alpha$</td>
<td>TSNR</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>GLOBAL</td>
<td>1.48</td>
<td>0.71</td>
<td>1.41</td>
<td>0.15</td>
</tr>
<tr>
<td>TROPICS</td>
<td>1.45</td>
<td>0.66</td>
<td>1.56</td>
<td>0.62</td>
</tr>
<tr>
<td>EXTROPICS</td>
<td>1.04</td>
<td>0.24</td>
<td>0.89</td>
<td>0.44</td>
</tr>
<tr>
<td>N HEM</td>
<td>1.26</td>
<td>1.03</td>
<td>1.27</td>
<td>0.25</td>
</tr>
<tr>
<td>S HEM</td>
<td>1.23</td>
<td>0.49</td>
<td>1.31</td>
<td>1.10</td>
</tr>
<tr>
<td>W HEM</td>
<td>1.07</td>
<td>0.03</td>
<td>1.38</td>
<td>0.39</td>
</tr>
<tr>
<td>E HEM</td>
<td>1.00</td>
<td>1.49</td>
<td>1.22</td>
<td>0.90</td>
</tr>
</tbody>
</table>

12.5.2 Solar signal

Detecting the solar signal was our original goal. Stevens and North (1996) found a fairly robust solar signal when the other deterministic signals were ignored. Taken as perpendicular to the individual signals it is still apparently detectable. For example, $\alpha \approx 0.5$, and TSNR $\approx 1.5$ when we use the component perpendicular to the volcanic vector. Similar results obtain when we look at the perpendicular to the greenhouse gas and the anthropogenic vectors. But when combined we obtain the rather poor performance shown in the table: TSNR $\approx 1.5, \alpha \approx 0.25$. This means our estimate of the solar cycle signal strength is $0.25 \pm 0.67$– too faint to detect with any confidence amidst this interference.

We made another attempt to estimate the solar signal by taking the 50 year data segment 1914-1963, a period in which there was very little volcanic activity. In the case that we considered the perpendicular component to all the other forcings we were able to obtain a value of $\alpha \sim 0.68$ with a TSNR of about 0.82 using EBM fluctuation data. Clearly, some performance was gained by considering this quite period but roughly a corresponding amount was lost because of the shorter record ($\sim \sqrt{2}$).
Table 12.6: Greenhouse signal estimated from its perpendicular component to the sum $G + A + V$. TSNR is a theoretical signal to noise ratio for a signal whose strength equals that of the EBM generated signal. The statistic $\alpha$ is the ratio of the estimated strength to that of the EBM generated signal. The SNR based upon the data is TSNR×$\alpha$.

<table>
<thead>
<tr>
<th></th>
<th>EBM</th>
<th></th>
<th>G/ML</th>
<th></th>
<th>G/C</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TSNR</td>
<td>$\alpha$</td>
<td>TSNR</td>
<td>$\alpha$</td>
<td>TSNR</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>GLOBAL</td>
<td>3.28</td>
<td>1.93</td>
<td>3.06</td>
<td>1.54</td>
<td>3.23</td>
<td>1.87</td>
</tr>
<tr>
<td>TROPICS</td>
<td>3.19</td>
<td>1.99</td>
<td>3.38</td>
<td>1.60</td>
<td>3.26</td>
<td>1.75</td>
</tr>
<tr>
<td>EXTROPICS</td>
<td>2.36</td>
<td>1.98</td>
<td>2.01</td>
<td>2.14</td>
<td>2.07</td>
<td>1.80</td>
</tr>
<tr>
<td>N HEM</td>
<td>2.86</td>
<td>1.94</td>
<td>2.83</td>
<td>1.78</td>
<td>2.74</td>
<td>1.53</td>
</tr>
<tr>
<td>S HEM</td>
<td>2.58</td>
<td>1.75</td>
<td>2.75</td>
<td>1.71</td>
<td>2.82</td>
<td>1.70</td>
</tr>
<tr>
<td>W HEM</td>
<td>2.49</td>
<td>2.03</td>
<td>3.13</td>
<td>1.73</td>
<td>2.92</td>
<td>1.96</td>
</tr>
<tr>
<td>E HEM</td>
<td>2.13</td>
<td>1.82</td>
<td>2.55</td>
<td>1.74</td>
<td>2.72</td>
<td>1.76</td>
</tr>
</tbody>
</table>

12.5.3 Greenhouse signal

The greenhouse signal is very strong in the data. We find SNR values consistently above 2.5 and values of $\alpha$ averaging about 1.80.

12.5.4 Anthropogenic aerosol signal

The statistical significance of the $A$ vector is poor if the value of $\alpha$ estimated were unity. If it truly had the strength the EBM suggests, it would have a very poor SNR. But the filtered results say the value is consistently much larger than the EBM calculated strength. This suggests that our SNR should be multiplied by its average estimated value of 3.70. The consistently large values of $\alpha$ throughout the table suggest that this is the proper interpretation. One explanation is that we have only included the direct influence of the aerosols and that the indirect (involving cloud particle generation) is actually several times larger.

12.5.5 Total deterministic signal

When we include all the signals added together and try to estimate the strength as detected in the noise we find a very strong signal (SNR $\geq 5.0$ for every entry in the table). The strength of the signal is $\alpha \approx 0.75 \pm 0.12$. 
Table 12.7: Anthropogenic aerosol signal estimated from its perpendicular component to the sum $G + S + V$. TSNR is a theoretical signal to noise ratio for a signal whose strength equals that of the EBM generated signal. The statistic $\alpha$ is the ratio of the estimated strength to that of the EBM generated signal. The SNR based upon the data is $\text{TSNR} \times \alpha$.

<table>
<thead>
<tr>
<th></th>
<th>EBM</th>
<th>G/ML</th>
<th>G/C</th>
<th>MPI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TSNR</td>
<td>$\alpha$</td>
<td>TSNR</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>GLOBAL</td>
<td>1.12</td>
<td>4.66</td>
<td>1.04</td>
<td>2.74</td>
</tr>
<tr>
<td>TROPICS</td>
<td>1.08</td>
<td>4.70</td>
<td>1.15</td>
<td>3.74</td>
</tr>
<tr>
<td>EXTROPICS</td>
<td>0.81</td>
<td>4.40</td>
<td>0.70</td>
<td>4.22</td>
</tr>
<tr>
<td>N HEM</td>
<td>0.99</td>
<td>4.40</td>
<td>0.97</td>
<td>3.31</td>
</tr>
<tr>
<td>S HEM</td>
<td>0.86</td>
<td>4.66</td>
<td>0.92</td>
<td>4.70</td>
</tr>
<tr>
<td>W HEM</td>
<td>0.85</td>
<td>4.40</td>
<td>1.03</td>
<td>4.09</td>
</tr>
<tr>
<td>E HEM</td>
<td>0.75</td>
<td>4.09</td>
<td>0.88</td>
<td>3.68</td>
</tr>
</tbody>
</table>

Table 12.8: Total deterministic signal $T = G + A + V + S$ as estimated from the data stream. TSNR is a theoretical signal to noise ratio for a signal whose strength equals that of the EBM generated signal. The statistic $\alpha$ is the ratio of the estimated strength to that of the EBM generated signal. The SNR based upon the data is $\text{TSNR} \times \alpha$.

<table>
<thead>
<tr>
<th></th>
<th>EBM</th>
<th>G/ML</th>
<th>G/C</th>
<th>MPI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TSNR</td>
<td>$\alpha$</td>
<td>TSNR</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>GLOBAL</td>
<td>8.33</td>
<td>0.82</td>
<td>7.25</td>
<td>0.64</td>
</tr>
<tr>
<td>TROPICS</td>
<td>8.01</td>
<td>0.85</td>
<td>8.41</td>
<td>0.88</td>
</tr>
<tr>
<td>EXTROPICS</td>
<td>5.58</td>
<td>0.91</td>
<td>5.03</td>
<td>0.75</td>
</tr>
<tr>
<td>N HEM</td>
<td>6.61</td>
<td>0.76</td>
<td>6.27</td>
<td>0.67</td>
</tr>
<tr>
<td>S HEM</td>
<td>7.07</td>
<td>0.89</td>
<td>7.51</td>
<td>0.97</td>
</tr>
<tr>
<td>W HEM</td>
<td>6.39</td>
<td>0.92</td>
<td>7.60</td>
<td>0.62</td>
</tr>
<tr>
<td>E HEM</td>
<td>5.39</td>
<td>0.93</td>
<td>6.00</td>
<td>0.69</td>
</tr>
</tbody>
</table>
12.6 Summary and Discussion

In this paper we presented a derivation of the optimal filter which is equivalent to earlier formulations, but conceptually much simpler because of its relation to the simple optimal weighting of independent estimators. A scheme was also developed for estimating individual signal strengths when they are embedded in a mixture of such deterministic signals. The method was applied to a series of signals which might be expected to appear in the data.

In applying the method we had to make use of a model to generate the expected waveform based upon some estimate of the forcing. In generating the signal waveform we used a simple energy balance model (Stevens, 1996; Stevens and North, 1996) which included seasonal variation and a high horizontal resolution (of order $10^\circ \times 10^\circ$).

In generating the optimal weights we used the GFDL and MPI coupled ocean/atmosphere climate models as well as the GFDL mixed layer climate model. We also used our noise forced EBM (Stevens, 1996). Because of the strong seasonal variation of the noise we kept only the summer half of each year in the middle latitudes and the whole year in the tropics. In this way, we were able to generate stationary time series in the control runs with one value per station per year. We made use of 36 stations in the analysis (Stevens and North, 1996) based upon record length, consistency, etc. The 36 component time series based upon the 4 control runs were then analyzed into frequency components and for each frequency component (there were 8) in a narrow band corresponding to periods ranging from 16.67 yr to 7.79 yr. Using only these relatively high frequency components we estimated strength of the following signals: solar ($S$), volcanic ($V$), greenhouse gas ($G$) and anthropogenic aerosol ($A$). Finally, we considered the total signal, $S + V + G + A$. The findings can be qualitatively summarized:

**Solar:** When the component of $S$ perpendicular to the other signals is considered so that no contamination occurs, we are not able to establish a solar signal with more than a few tens of percent of confidence. This does not mean it is not there, but with the data we used combined with our method, it is too faint to be established firmly. We will continue to look for this elusive signal in other data sets (e.g., longer records involving fewer stations, etc.). One important finding in this analysis is the chance occurrence of the several climatically significant volcanic eruptions in the last 40 yr. These eruptions have a spacing of about 10 years and could be seriously contaminating some estimates of solar response when only a few decades of data are used in the search. Similarly spectral approaches such as ours run the risk of
significant leakage through the filter of greenhouse and anthropogenic aerosol forcings in the solar band. While the band is quite narrow, the strengths of $G$ and $A$ are so large as to potentially contaminate the estimation.

**Volcanic:** The volcanic signal estimated by taking its component perpendicular to all other signals has a strength about little larger than but consistent with its EBM predicted strength. The theoretical signal to noise ratio for this estimate is of the order of 4.0. We believe this firmly establishes the signature of the volcanic response in the record.

**Greenhouse gas:** This signal is very strong when taken perpendicular to the others. Its theoretical signal-to-noise ratio is over 3.0 for all models and its estimated amplitude is larger than the EBM calculated value by at least 50% over all models.

**Anthropogenic aerosols:** The anthropogenic signal alone is very strong ($\sim 2$ to 4 times the EBM calculated amount), but its TSNR is only of the order of unity. One interpretation is that it really is that strong with a correspondingly large SNR, but we are cautious about taking this position. One interpretation is that both $G$ and $A$ are under calculated by the EBM and the cancellation is very strong.

$S + V + G + A$, the sum of all signals. In this case we have a signal close to that calculated by the EBM (0.64 to 0.82, depending upon the model used in the filter construction), but the theoretical signal-to-noise ratio is between 7.0 and 11! We draw the same conclusion as other recent studies (Santer et al., 1996, Hegerl et al., 1996), that as we combine the signals we find overwhelming evidence of their presence in the data stream.

### 1 Errors Due to Incorrect EOFs

In the early sections of this paper we assumed that we could rotate our coordinate system to the one in which the $n_i$ were mutually uncorrelated. In practice this transformation is difficult, since there is no data available for the purpose of finding the covariance matrix of the undisturbed climate. The climate record as we know it is a combination of natural variability and forced responses to various stimuli. The only recourse is to use the control runs of numerical models of the system to obtain the covariance statistics of the undisturbed climate. In the optimal weighting scheme two important ingredients come from the determination of the EOFs: the eigenvalues of the covariance matrix $\lambda_i$, and the direction cosines of the signal, $e_S \cdot e_i$. 

.1. ERRORS DUE TO INCORRECT EOFs

To illustrate the problem consider the simple case of the two unbiased thermometers $\hat{T}_1, \hat{T}_2$ with different error variances $\sigma_1^2, \sigma_2^2$. As was found in Section 2, we find that the MSE is a quadratic function of the weight coefficient $A$. As a numerical illustration Fig. A1 shows the MSE for $\sigma_1 = 3, \sigma_2 = 4$. The broad minimum in the curve suggests that a slightly incorrect choice of $A$ leads to a negligible change in the MSE. It also shows that for this choice of variances, there is considerable gain in using the optimal weighting over ignoring the high variance data or over a straight numerical average of the data.

Now suppose that our assumption that the correlation between the errors in the two instruments is incorrect. As an example, suppose the actual correlation between the two instruments is $\rho = 0.3$. If this were true the actual least MSE would be about 7.5 instead of 6.0, a substantial increase. We would have found the coefficient $A$ to be 0.61 instead of about 0.70. The erroneous choice of $A$ leads to an increase of the actual MSE of only about 0.4 or about 6%. Hence, the MSE is only increased marginally, but our estimate of the MSE is about 25% too low. Using the wrong EOFs (coordinate system) in this case does not lead to a substantially increased error, but it does c
Appendix A

Estimating Space-Time Averages

Suppose we wish to estimate the area average of a field, but our observations are limited to a point or two within the area. Consider the domain \( \mathcal{D} \) in which there is a point measurement taken at the point \( P, \hat{T}_P \). We are taking the point measurement as an estimate of the area average of \( T(r) \) over the domain \( \mathcal{D} \). The error is the difference between the estimate and the true value

\[
\text{Error} = \frac{1}{\mathcal{D}} \int_{\mathcal{D}} (1 - D\delta(r - r_P))T(r) \, d^2r
\]  

(A.1)

The MSE is given by

\[
\text{MSE} = \langle \text{Error}^2 \rangle = \frac{1}{D^2} \int_{\mathcal{D}} \int_{\mathcal{D}} (1 - D\delta(r - r_P))(1 - D\delta(r' - r_P))\langle T(r)T(r') \rangle \, d^2r \, d^2r'
\]  

(A.2)

To proceed consider the case of a homogeneous field; i.e.,

\[
\langle T(r) \rangle = 0,
\]

\[
\langle T(r)T(r') \rangle = C(|r - r'|)
\]  

(A.4)

(A.5)

This allows us to use the Fourier spectral form

\[
C(|r - r'|) = \int \int S(\nu)e^{2\pi i\nu \cdot (r - r')} \, d^2r \, d^2r'
\]  

(A.6)
The integrals in the last formula are to be taken over all space. We are left with the neat form

\[ \epsilon^2 = \int |H(\vec{\nu})|^2 S(\nu) \, d^2 \vec{\nu} \]  
(A.7)

where

\[ H(\vec{\nu}) = \frac{1}{D} \int_D (1 - D\delta(\vec{r} - \vec{r}_P)) e^{2\pi i \vec{\nu} \cdot \vec{r}} \, d^2 \vec{r} \]  
(A.8)

As an example consider a square whose edge has length \( L \). Then

\[ H(\vec{\nu}) = G(\pi L \nu_x)G(\pi L \nu_y) - e^{2\pi i \vec{\nu} \cdot \vec{r}_P} \]  
(A.9)

where

\[ G(z) \equiv \frac{\sin z}{z} \]  
(A.10)

Hence,

\[ |H(\vec{\nu})|^2 = (G_x G_y - \cos(2\pi \vec{\nu} \cdot \vec{r}_P))^2 + \sin^2(2\pi \vec{\nu} \cdot \vec{r}_P) \]  
(A.11)

\[ = G_x^2 G_y^2 - 2G_x G_y \cos(2\pi \vec{\nu} \cdot \vec{r}_P) + 1 \]  
(A.12)

It is not difficult to see that the MSE is least if the point \( \vec{r}_P \) is at the center of the square.

The formula (A.7) is very powerful. For example, can consider the case of \( N \) weighted point measurements inside the domain \( D \). Then

\[ H(\vec{\nu}) = G_x G_y - \sum_n \alpha_n e^{2\pi i \vec{\nu} \cdot \vec{r}_n} \]  
(A.13)

\[ = \sum_n \alpha_n \{G_x G_y - e^{2\pi i \vec{\nu} \cdot \vec{r}_n}\} \]  
(A.14)

\[ = \sum_n \alpha_n H_n(\vec{\nu}) \]  
(A.15)

We may next consider the magnitude squared:

\[ |H(\vec{\nu})|^2 = \sum_n \alpha_n^2 |H_n(\vec{\nu})|^2 + \sum_{n,m} \alpha_n \alpha_m H_n(\vec{\nu}) H_m^*(\vec{\nu}) \]  
(A.16)

To find the optimal weights we minimize \( \epsilon^2 \) wrt the \( \alpha_i \) subject to the condition \( \sum_i \alpha_1 = 1 \). The latter can usually best be taken into account with the use of a Lagrange multiplier.
Appendix B

Note on Statistical Extrapolation of Fields

Linear statistical extrapolation or prediction is not new. Here we adapt an old technique to the problem in meteorology and climate: how do we make a purely linear statistical forecast of information in the future along the time axis or onto some neighboring spatial domain, where we do not have data? The new aspect is the fact that our application is to fields which are continuous in space and evolving in time. We seek a formulation that makes use of the eigenfunctions of the covariance kernels. We believe this is fundamental in climate dynamics. The complete solution will be given in terms of the space-time EOFs on the whole space-time domain and the space-time EOFs on the restricted data domain (the subregion of space-time where predictor information is available).

B.1 Optimal Filter Formulation

We have a space-time domain $R$ in which a field $\psi(x)$ is defined. We wish to predict $\psi(x)$ at a point $x \in R$. The prediction is to be based upon known values of $\psi(x)$ in a subset of $R$, namely $D$, $D \subset R$. The prediction is to be optimal in the least MSE sense. Let $\hat{\psi}(x)$ be the predicted value at point $x$. Then our predictor can be defined as

$$\hat{\psi}(x) = \int_D \Gamma(x, x') \psi(x') \, dx' \quad (B.1)$$
\( \int_R \chi(x') \Gamma(x, x') \psi(x') \, dx' \)  

where

\[
\chi(x) = \begin{cases} 
1, & x \in D \\
0, & \text{otherwise} 
\end{cases} 
\]

and \( \Gamma(x, x') \) is a filter which we hope to tailor to our needs. The filter provides a weighting over the data that leads to a prediction. The MSE is defined as

\[
\epsilon^2 = \langle (\psi(x) - \hat{\psi}(x))^2 \rangle 
\]

where \( \langle \cdot \rangle \) means ensemble average. Our goal is to minimize \( \epsilon^2 \) by adjusting the shape of \( \Gamma(x, x') \).

We take the variation of \( \epsilon^2 \) and set it to zero:

\[
- \left\langle \psi(x) \int_R \chi(x') \delta \Gamma(x, x') \psi(x') \, dx' \right\rangle 
+ \left\langle \int_R \Gamma(x, x'') \chi(x''') \psi(x''') \, dx''' \int_R \delta \Gamma(x, x') \chi(x') \psi(x') \, dx' \right\rangle = 0 \]  

As usual if this is to be true for arbitrary \( \delta \Gamma \), we have

\[
\int_R \Gamma(x, x'') \chi(x''') C(x''', x') \, dx''' = \chi(x') C(x, x') 
\]

where the covariance kernel is defined by:

\[
C(x, x') \equiv \langle \psi(x) \psi(x') \rangle 
\]

The solution of this integral equation is the optimal filter \( \Gamma^{opt}(x, x') \). Note that because of the presence of \( \chi(x''') \) under the integral sign, we have some difficulties.

It might be convenient to use the known basis set defined on \( R \):

\[
\int_R C(x, x') \phi_n(x') \, dx' = \gamma_n \phi_n(x) 
\]

then

\[
C(x, x') = \sum_n \gamma_n \phi_n(x) \phi_n(x') 
\]

\[
\Gamma(x, x') = \sum_n \Gamma_n(x) \phi_n(x') 
\]

We also have completeness

\[
\sum_n \phi_n(x) \phi_n(x') = \delta(x - x') 
\]
B.1. OPTIMAL FILTER FORMULATION

Note that these eigenfunctions are defined on the whole domain \( R \), but the integral in (B.6) only runs over the subdomain, \( D \).

Now we can use the orthogonal basis set \( \zeta_n(x) \) defined on the subregion \( D \)

\[ \int_D C(x, x') \zeta_n(x') \, dx' = \gamma_n' \zeta_n(x) \quad (B.12) \]

Then

\[ \sum_n \Gamma_n(x) \gamma_n' \zeta_n(x') = \chi(x') C(x, x') \quad (B.13) \]

Where the superscript \( \zeta \) on \( \Gamma_n(x) \) indicates that the projection is with respect to the \( \zeta_n \) basis set. Next we project the \( n \)th component

\[ \Gamma_n(x) \gamma_n' = \int_D C(x, x') \zeta_n(x') \, dx' \quad (B.14) \]

and dividing thru by \( \gamma_n' \), multiplying by \( \zeta_n(x) \) and summing

\[ \Gamma^{opt}(x, x') = \sum_n \frac{1}{\gamma_n} \int_D C(x, x') \zeta_n(x) \zeta_n(x') \, dx' \quad (B.15) \]

Finally, it might be useful to write the \( C(x, x') \) in terms of the \( \phi_n(x) \), using (B.9):

\[ \Gamma^{opt}(x, x') = \sum_{nm} \frac{\gamma_n}{\gamma_m} \phi_m(x) \zeta_n(x') (\phi_m, \zeta_n)_D \quad (B.16) \]

where we used the inner product notation:

\[ (\phi_m, \zeta_n)_D \equiv \int_D \phi_m(x') \zeta_n(x') \, dx' \quad (B.17) \]

Note that \( \phi_m(x) \) is defined for values of \( x \) outside of \( D \), which allows \( \Gamma^{opt}(x, x') \) to be defined for \( x \) throughout \( R \). The last form can also be written:

The above suggests that the optimal linear predictor is easily found if the functions \( \phi_m(x), \zeta_n(x) \) and the eigenvalues \( \gamma_n, \gamma'_n; n, m = 0, 1, 2, \ldots \) are found in advance.

Note that if \( x \in D \), the last formula collapses: \( \gamma_n = \gamma_n' \), \( \phi_n(x) = \zeta_n(x) \), and finally \( \Gamma(x, x') = \delta(x - x') \), where we used (B.11).

The final formula for the prediction is

\[ \hat{\psi}(x) = \sum_{nm} \frac{\gamma_n}{\gamma_m} \phi_m(x) (\psi, \zeta_n)_D (\zeta_n, \phi_m)_D \quad (B.18) \]
APPENDIX B. NOTE ON STATISTICAL EXTRAPOLATION OF FIELDS

B.2 Examples (incomplete)

B.2.1 One Dimensional Interpolation

Consider a homogeneous random function (equivalent to a stationary time series), \( T(x) \). We have data at discrete points that are equally spaced, \( x_n = n \Delta x, n = 0, \pm 1, \pm 2, \ldots \); these \( x_n \) constitute the data set, \( D \). The \( \phi_\nu(x) \) are the Fourier integral components:

\[
\phi_\nu(x) = e^{2\pi i \nu x}
\]

To find the \( \zeta_n(x_i) \), we must solve

\[
\sum_n K(m,n)\zeta_l(x_n) = \gamma' \zeta_l(x_m) \tag{B.20}
\]

where

\[
K(m,n) = \langle T(x_m)T(x_n) \rangle = C((m - n)\Delta x) \tag{B.21}
\]

Next set

\[
C((m - n)\Delta x) = \int_{-L/2}^{L/2} \tilde{C}_\nu e^{2\pi i \nu (m-n)\Delta x} \tag{B.22}
\]

where \( L = 1/\Delta x \). This leaves

\[
\int d\nu \sum_n \tilde{C}_\nu e^{2\pi i \nu m \Delta x} e^{-2\pi i \nu n \Delta x} \zeta(n\Delta x) = \gamma \zeta(m\Delta x). \tag{B.23}
\]

Now define

\[
\tilde{\zeta}(\nu) = \sum_n \zeta(n\Delta x)e^{2\pi i \nu n \Delta x} \tag{B.24}
\]

Then

\[
\int \tilde{C}_\nu e^{2\pi i \nu m \Delta x} \tilde{\zeta}(-\nu) = \gamma \zeta(m\Delta x) \tag{B.25}
\]

Now multiply through by \( e^{2\pi i \nu' m \Delta x} \) and perform the sum over \( m \), noting that \( \sum_m e^{2\pi i (\nu - \nu') \Delta x m} = \delta((\nu - \nu')\Delta x) \)

\[
\int d\nu \tilde{C}_\nu \delta(\nu - \nu') \tilde{\zeta}(-\nu) = \gamma \tilde{\zeta}(-\nu') \tag{B.26}
\]

This tells us that the (continuum) eigenvalues are

\[
\gamma = \tilde{C}_\nu \tag{B.27}
\]

and the basis functions are the discrete Fourier basis

\[
\zeta_\nu(n\Delta x) = e^{2\pi i \nu n \Delta x} \tag{B.28}
\]
B.2.2 Prediction

Suppose we have a stationary time series on $-\infty < t < \infty$ and we have data for the subdomain, $D$ defined by the interval $-T/2 < t < T/2$. The $\phi_f(t)$ are the Fourier integral basis and the $\zeta_\lambda(t)$ are the eigenfunctions:

$$\int_D C(t-t')\zeta_\lambda(t')\,dt' = \sigma_\lambda \zeta_\lambda(t) \quad (B.29)$$

The functions $\zeta_\lambda(t)$ will depend on the spectrum of the underlying process on $\mathbb{R}$. Use

$$C(\tau) = \int_{-\infty}^{\infty} S(f)e^{2\pi if\tau}\,df \quad (B.30)$$

Next define

$$\tilde{\zeta}_\lambda(f) \equiv \int_D \zeta_\lambda(t)e^{2\pi ift}\,dt \quad (B.32)$$

Multiplying through (B.31) by $e^{2\pi if't}$ and integrating over $D$ leads to a new eigenvalue problem:

$$\int_{-\infty}^{\infty} df S(f)\tilde{\zeta}_\lambda T(f)T G_T(f-f') = \sigma_\lambda \tilde{\zeta}_\lambda T(f') \quad (B.33)$$

where we have used

$$G_T(f) \equiv \frac{\sin(\pi fT)}{\pi fT} \quad (B.34)$$

In the last integral equation it is better to do some scaling to symmetrize the integral equation: define

$$\tilde{\xi}_{\lambda T}(f) \equiv \sqrt{S(f)}\tilde{\zeta}_{\lambda T}(f)$$

and

$$G'_T(f,f') = \sqrt{S(f)}G_T(f-f')\sqrt{S(f')}$$

, then

$$\int_{-\infty}^{\infty} G'_T(f,f')\tilde{\xi}_{\lambda T}(f')\,df' = \sigma_\lambda \tilde{\xi}_{\lambda T}(f) \quad (B.35)$$

Hence, to solve the extrapolation problem, we need to first find the eigenfunctions in the above system and then we can construct the predictor.
APPENDIX B. NOTE ON STATISTICAL EXTRAPOLATION OF FIELDS
Appendix C

Principal Oscillation Patterns

The idea is to represent a time dependent stream of data with a simple linear stochastic model. When this is achieved, the model can be used to interpret the underlying processes. The original idea is due to Hasselmann. A vector notation is introduced with bold indicating vectors, and caligraphic indicating matrices. The process is to follow the model:

\[ x(t+1) = A \cdot x(t) \]  

(C.1)

with all entries real. If the matrix \( A \) were symmetric the eigenvalues would be real and the eigenvectors would be orthogonal. Here we relax the symmetry condition. Then the eigenvectors may be defined by

\[ A \cdot p = \lambda p \]  

(C.2)

Both \( \lambda \) and \( p \) might be complex. It is important to note though that since \( A \) is real we have:

\[ A \cdot p^* = \lambda^* p^* \]  

(C.3)

In other words, if \( p \) is an eigenvector, then so is \( p^* \) and it corresponds to eigenvalue \( \lambda^* \). The vectors \( p \) are known as the principal oscillation patterns (POPs).

The state \( x \) at any time \( t \) can be expressed in terms of the \( p \):

\[ x = \sum_j z_j(t) p_j \]  

(C.4)

Then substituting:

\[ z(t+1) \cdot p = \lambda \cdot z(t) \cdot p \]  

(C.5)
If \( z(0) = 1 \), then
\[
z(t) \cdot p = \lambda^t \cdot p \tag{C.6}
\]
The contribution \( P(t) \) of the complex conjugate pair \( p, p^* \) is given by
\[
P(t) = z(t) \cdot p + [z(t) \cdot p]^* \tag{C.7}
\]
Let
\[
p = p^r + i \cdot p^i \tag{C.8}
\]
\[
z(t) = \frac{1}{2}(z^r - i \cdot z^i) \tag{C.9}
\]
Then
\[
P(t) = z^r(t) \cdot p^r + z^i(t) \cdot p^i \tag{C.10}
\]
\[
= \rho^t \cos(\eta t) \cdot p - \sin(\eta t) \cdot p^i \tag{C.11}
\]
with \( \lambda = \rho \exp(-i\eta) \). The system trajectory spirals into the origin with period \( 2\pi/\eta \) and decay time \( -1/\ln \rho \).

The modes may be represented by relative amplitude and phase:
\[
A^2(r) = [p^r]^2 + [p^i]^2 \tag{C.12}
\]
\[
\psi(r) = \tan^{-1}[p^i]/[p^r] \tag{C.13}
\]

### C.1 Adjoint Patterns

The adjoint patterns \( p^A \) are the eigenvectors of \( A^T \). These eigenvectors are orthogonal to the \( p, (p^A)^T \cdot p_k = \delta_{jk} \). The pattern coefficients can be calculated from
\[
(p^A)^T x = \sum_k z_k (p^A)^T p_k = z_j \tag{C.14}
\]
The eigenvalues for the adjoint patterns are the same as for the regular patterns.

### C.2 Stochastic Model

The associated stochastic model is
\[
x(t + 1) = A \cdot x(t) + \text{noise} \tag{C.15}
\]
Multiply by \( x^T(t) \) and take expectations leads to:
\[
A = \langle [x(t + 1)x^T(t)] \rangle \cdot \langle [x(t)x^T(t)] \rangle^{-1} \tag{C.16}
\]
C.3. Estimating the System Matrix

If we have a stream of vector data $\mathbf{x}(t)$, we can estimate the system matrix $A$ by a least mean square error technique. First form the mean square error:

$$
e^2 = \langle \sum_i (x_i(t+1) - A_{ij}x_j(t))^2 \rangle$$  \hspace{1cm} (C.17)

$$= \langle \left(\sum_i x_i(t+1)\right)^2 - 2 \sum_{ij} A_{ij} \langle x_i(t+1)x_j(t) \rangle \rangle$$  \hspace{1cm} (C.18)

$$+ \sum_{ijk} A_{ij}A_{ik} \langle x_j(t)x_k(t) \rangle$$  \hspace{1cm} (C.19)

Now taking partial derivatives:

$$\frac{\partial e^2}{\partial A_{il}} = -2\langle x_i(t+1)x_l(t) \rangle + 2 \sum_j A_{ij} \langle x_j(t)x_l(t) \rangle$$  \hspace{1cm} (C.20)

$$= 0$$  \hspace{1cm} (C.21)

We finally arrive at

$$A_{ij}^{opt} = \sum_j D_{ij} (M^{-1})_{jl}$$  \hspace{1cm} (C.22)

where

$$M_{il} = \langle x_i(t)x_l(t) \rangle$$  \hspace{1cm} (C.23)

$$D_{jl} = \langle x_j(t+1)x_l(t) \rangle$$  \hspace{1cm} (C.24)

C.4. Cyclostationary POPs

Time is given by a pair of integers $(t, \tau)$ with $t$ counting the cycles and $\tau$ counting the seasonal date. If $n$ is the number of steps in a cycle, $(t, \tau+n) = (t+1, \tau)$. The cyclostationary process is:

$$\mathbf{x}(t, \tau+1) = A(\tau) \cdot \mathbf{x}(t, \tau)$$  \hspace{1cm} (C.25)

with

$$\mathbf{x}(t, \tau+n) = \mathbf{x}(t+1, \tau)$$  \hspace{1cm} (C.26)

$$A(\tau+n) = A(\tau)$$  \hspace{1cm} (C.27)

Applying the advance operator $n$ times we have

$$B(\tau) = A(\tau+n-1) \cdots A(\tau+1)A(\tau)$$  \hspace{1cm} (C.28)
so that

\[ x(t + 1, \tau) = B(\tau)x(t, \tau) + \text{noise} \quad (C.29) \]

For each \( \tau \) we can now form eigenvectors (patterns):

\[ B(\tau)p^\tau = \lambda^\tau p^\tau \quad (C.30) \]

The eigenvalues are the same for the different \( B(\tau) \) models as seen by

\[ B(\tau + n)A(\tau + n)p^\tau = \lambda^\tau A(\tau + n)p^\tau \quad (C.32) \]

\[ B(\tau + 1)[A(\tau)p^\tau] = \lambda^\tau [A(\tau)p^\tau] \quad (C.33) \]

Thus \( B(\tau + 1) \) and \( B(\tau) \) share the same eigenvalues and \( A(\tau)p^\tau \) is an eigenvector of \( B(\tau + 1) \) if \( p^\tau \) is an eigenvector of \( B(\tau) \).

Eigenvectors may be multiplied by any complex number \( c = r^{-1} \exp i\phi \), (see von Storch et al.). We eventually get

\[ z(t, \tau + 1) = r_\tau \exp(-i\eta/n)z(t, \tau) + \text{noise} \quad (C.34) \]

where the eigenvalue \( \lambda = \rho \exp(-i\eta) \).

Appendix D

Singular Value Decomposition

This is a method used by some investigators in correlating one field with another. There are several ways of doing this as summarized by Bretherton et al. (1992) (hereafter referred to as BSW). Basically we have two fields $\Psi(x)$ and $\Phi(x)$, where $x = (r, t)$ is a point in space-time. The fields are to be random. That is we are to consider realizations of the fields in space-time. There may be autocorrelations in the fields. In some sense the fields are correlated with one another. We ask here about some ways of bringing out that correlation such that there is some dynamical meaning to the correspondence. We differ from BSW in that we use continuous fields rather than a grid point representation. This is merely a matter of notation.

D.1 Covariance

We consider the raw covariance kernel

$$C(x, x') \equiv \langle \Psi(x) \Phi(x') \rangle$$  \hspace{1cm} (D.1)

where it is assumed that the means have been removed:

$$\langle \Psi(x) \rangle = \langle \Phi(x) \rangle = 0$$  \hspace{1cm} (D.2)

The transpose of the covariance kernel is

$$C^T(x, x') \equiv \langle \Phi(x) \Psi(x') \rangle$$  \hspace{1cm} (D.3)
According to BSW the covariance can be decomposed uniquely by
\[ C(x, x') = \sum_k \sigma_k \psi_k(x) \phi_k(x') \] (D.4)
and
\[ C^T(x, x') = \sum_k \sigma_k \psi_k(x') \phi_k(x) \] (D.5)
where the \( \psi_k(x) \) and \( \phi_k(x) \) are each complete sets of orthonormal functions:
\[ \int \phi_n(x) \phi_m(x) = \delta_{nm} \] (D.6)
\[ \int \psi_n(x) \psi_m(x) = \delta_{nm} \] (D.7)
Using the decomposition (D.4) we can easily establish that
\[ \int C(x, x') \phi_k(x) \, dx = \sigma_k \psi_k(x) \] (D.8)
\[ \int C^T(x, x') \psi_k(x) \, dx = \sigma_k \phi_k(x) \] (D.9)
Next consider the matrix product of the kernel with its transpose
\[ CC^T = \int C(x, x'') C^T(x'', x') \, dx'' \] (D.10)
\[ = \sum_{k,k'} \sigma_k \sigma_{k'} \int \psi_k(x) \phi_k(x'') \psi_{k'}(x'') \phi_{k'}(x') \, dx'' \] (D.11)
\[ = \sum_k \sigma_k^2 \psi_k(x) \psi_k(x') \] (D.12)
Similarly,
\[ C^T C = \sum_k \sigma_k^2 \phi_k(x) \phi_k(x') \] (D.13)
We note that both \( CC^T \) and \( C^T C \) are symmetric kernels (matrices) and therefore they have orthogonal eigenfunctions (vectors) with positive eigenvalues. In fact, we can see from the last two equations that the eigenfunctions of \( CC^T \) are the \( \psi_k(x) \) with eigenvalues \( \sigma_k^2 \) and the eigenfunctions of \( C^T C \) are the \( \phi_k(x) \) also with eigenvalues \( \sigma_k^2 \). These last two statements allow us a method of computing the \( \psi_k(x) \) and the \( \phi_k(x) \) by straightforward algorithms using only the known quantities \( C \) and \( C^T \). Any function can now be expanded into these orthonormal sets of functions. Note, however, that the \( \psi_k(x) \) are not necessarily orthogonal to the \( \phi_n(x) \).
We can expand an individual realization of the fields:

\[
\Psi(x) = \sum_k a_k \psi_k(x) \quad \text{(D.14)}
\]

\[
\Phi(x) = \sum_k a_k \phi_k(x) \quad \text{(D.15)}
\]

Since the fields are random and the eigenfunctions are not, the coefficients \(a_k\) and \(b_k\) are random variables, changing from one realization to another. The coefficients may be computed from

\[
a_k = \int \psi_k(x) \Psi(x) \, dx \quad \text{(D.16)}
\]

\[
b_k = \int \phi_k(x) \Phi(x) \, dx \quad \text{(D.17)}
\]

Using these we can compute

\[
\langle a_k b_{k'} \rangle = \int \int \psi_k(x) \phi_{k'}(x') \langle \Psi(x) \Phi(x') \rangle \, dx \, dx' \quad \text{(D.18)}
\]

\[
= \sum_n \sigma_k \delta_{kn} \delta_{k'n} \quad \text{(D.19)}
\]

\[
= \sigma_k \delta_{kk'} \quad \text{(D.20)}
\]

Hence, the coefficients in these expansions are uncorrelated except for the matching indices. Thus the expansions are such that each field is expanded into “modes” and the modes of one are corresponded to those of the other. It can be shown that this is the most economical choice in the sense that if only \(N\) such terms are retained, the most variance will be captured by these expansion functions (over any other choice).