Chapter 7
Time Series Analysis

7.1 Introduction

In this chapter we will consider some common aspects of time series analysis including autocorrelation, statistical prediction, harmonic analysis, power spectrum analysis, and cross-spectrum analysis. We will also consider space-time cross spectral analysis, a combination of time-Fourier and space-Fourier analysis, which is often used in meteorology. The techniques of time series analysis described here are frequently encountered in all of geoscience and in many other fields. We will spend most of our time on classical Fourier spectral analysis, but will mention briefly other approaches such as Maximum Entropy (MEM), Singular Spectrum Analysis (SSA) and the Multi-Taper Method (MTM). Although we include a discussion of the historical Lag-correlation spectral analysis method, we will focus primarily on the Fast Fourier Transform (FFT) approach.

7.2 Autocorrelation and Red Noise

7.2.1 The Autocorrelation Function

Given a continuous function $x(t)$, defined in the interval $t_1 < t < t_2$, the autocovariance function is

$$\Phi(\tau) = \frac{1}{t_2 - t_1 - \tau} \int_{t_1}^{t_2-\tau} x'(t)x'(t+\tau)dt \tag{7.1}$$

where primes indicate deviations from the mean value, and we have assumed that $\tau > 0$. In the discrete case where x is defined at N points spaced at an interval of Δt , $k = 1, 2, .., N$, we can calculate the autocovariance at lag L.

$$\Phi(L) = \frac{1}{N - 2L} \sum_{k=L}^{N-L} x'_k x'_{k+L} = \overline{x'(t)x'(t + L\Delta t)} \tag{7.2}$$

The autocovariance is the covariance of a variable with itself (Greek autos = self) at some other time, measured by a time lag (or lead) τ. Note that $\Phi(0) = \overline{x'^2}$, so that the autocovariance at lag zero is just the variance of the variable.

The Autocorrelation function is the normalized autocovariance function $r(\tau) = \Phi(\tau)/\Phi(0)$ and $-1 < r(\tau) < 1$. If x is not periodic then $r(\tau) \to 0$, as $\tau \to \infty$ It is normally assumed that data sets subjected to time series analysis are stationary . The term stationary time series normally implies that the true mean of the variable and its higher-order statistical moments are independent of the particular time in question and so do not vary within the sample. Therefore it is usually necessary to remove any trends in the time

series before analysis. This also implies that the autocorrelation function can be assumed to be symmetric, $r(\tau) = r(-\tau)$.

7.2.2 White Noise

In the special case $r(\Delta t) = a = 0$, our time series is a series of random numbers, uncorrelated in time so that $r(\tau) = \delta(0)$ a delta function. For such a "white noise" time series, the present value is of no help in projecting into the future. The probability density function of white noise can vary, but we will generally use Gaussian Normal white noise whose probability distribution is Gaussian around a mean value of zero. Figure 7.1 shows a sample of Gaussian and uniformly distributed white noise, along with their corresponding sample probability density functions. Gaussian noise is both more likely to be near zero and to depart far from zero than uniformly distributed noise. A Gaussian distribution fits many naturally occurring time series. In both cases the autocorrelation is zero for any nonzero lag.

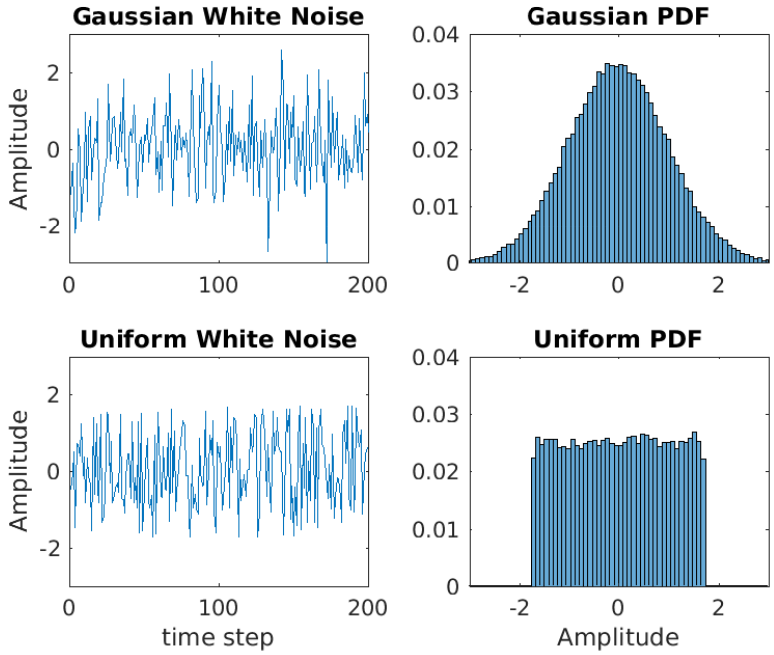

Figure 7.1 Samples of Gaussian and Uniform White Noise and their sample probability distributions. Both samples have a standard deviation of 1.0. The probability distributions are based on a sample of 50,000 time steps.

7.2.3 Red Noise

We define a "red noise" time series as being of the form:

$$x(t) = ax(t - \Delta t) + \left(1 - a^2\right)^{1/2} \epsilon(t) \tag{7.3}$$

where x is a standardized variable $\overline{x} = 0$ and $\overline{x'^2} = 1$. The constant a is on the interval between zero and one $(0 < a < 1)$ and measures the degree to which memory of previous states is retained. The function ϵ represents a series of random numbers drawn from a standardized normal distribution, and Δt is the time

interval between data points. This is also called a Markov Process or an Auto-Regressive, or AR-1 Process, since it remembers only the previous value.

Multiply 7.3 by $x(t-\Delta t)$ and average over time to show that a is the one-lag autocorrelation, or the autocorrelation at one time step Δt.

$$\overline{x(t)x(t-\Delta t)} = a\overline{x(t-\Delta t)x(t-\Delta t)} + \left(1-a^2\right)^{1/2}\overline{\epsilon(t)x(t-\Delta t)} \tag{7.4}$$

Since the time series has unit variance, and the noise is uncorrelated with the time series, 7.4 becomes,

$$r(\Delta t) = a \tag{7.5}$$

so that a is the autocorrelation at one time step.

Using 7.4 multiple times we can show how the autocorrelation depends on the time interval.

$$\begin{aligned} x(t+\Delta t) &= ax(t) + \left(1-a^2\right)^{1/2}\epsilon(t) \\ x(t+\Delta t) &= a(ax(t-\Delta t) + \left(1-a^2\right)^{1/2}\epsilon(t)) + \left(1-a^2\right)^{1/2}\epsilon(t) \\ \overline{x(t+\Delta t)x(t-\Delta t)} &= a^2\overline{x(t-\Delta t)x(t-\Delta t)} + 0 \\ r(2\Delta t) &= a^2 \end{aligned} \tag{7.6}$$

From 7.6 we determine by induction that,

$$r(n\Delta t) = a^n \tag{7.7}$$

So for a red noise time series, the autocorrelation at a lag of n time steps is equal to the autocorrelation at one lag, raised to the power n. A function that has this property is the exponential function, $e^{nx} = \left(e^x\right)^n$, so we may hypothesize that the autocorrelation function for red noise has an exponential shape.

$$r(n\Delta t) = \exp\left(-n\Delta t/T\right) \tag{7.8}$$

where $T = -\Delta t/\ln a$ is the e-folding time of the autocorrelation, and if $\tau = n\Delta t$, then

$$r(\tau) = \exp\left(-|\tau|/T\right) \tag{7.9}$$

The autocorrelation function 7.9 is shown in Fig. 7.2.

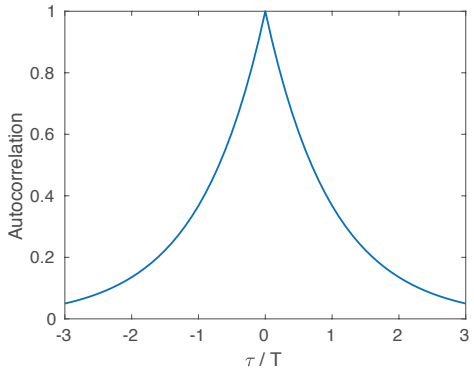

Figure 7.2 Autocorrelation function for red noise 7.9 plotted as a function of lag time τ divided by e=folding time T.

In summary, if we are given an auto-regressive (AR-1) process 7.3, then its autocorrelation is given by 7.9, where the e-folding time is $T = -\Delta t/\ln a$.

7.3 Statistical Prediction and Red Noise

Consider a prediction equation of the form

$$\widehat{x}(t+\Delta t) = a_1 x(t) + a_2 x(t-\Delta t) \tag{7.10}$$

where a_1 and a_2 are chosen to minimize the root-mean-square error on dependent data. Recall from our discussion of multiple regression that for two predictors x_1 and x_2 used to predict y, the second predictor is only useful if,

$$\left|r(x_2, y)\right| \geqslant \left|r(x_1, y) r(x_1, x_2)\right| \tag{7.11}$$

In the case where the equality holds, $r(x_2, y)$ is equal to the "minimum useful correlation" discussed in 4.3.3 and will not improve the forecasting skill beyond the level possible by using x_1 alone. In the case of trying to predict future values from prior times, $r(x_2, y) = r(2\Delta t)$, and $r(x_1, y) = r(x_1, x_2) = r(\Delta t)$ so that we must have,

$$\left|r(2\Delta t)\right| \geqslant r(\Delta t)^2 \tag{7.12}$$

in order to justify using a second predictor at two time steps in the past. Note that for red noise

$$r(2\Delta t) = r(\Delta t)^2 \tag{7.13}$$

so that the value at two lags previous to now always contributes exactly the minimum useful, and nearly automatic, correlation, and not more. For red noise, then, nothing is gained by using a second predictor. All we can use productively is the present value and the autocorrelation function. Using one predictor one time step before is called the persistence forecast, where we assume tomorrow will be like today.

7.4 Degrees of Freedom with Gaussian Red Noise

For a red noise process, adjacent values are correlated and so not independent. One cannot gain more information about a red noise process by sampling it at finer and finer temporal resolution. Information, or degrees of freedom, increase as a function of sample length. Leith (1973) used a Gaussian red noise model to assess the number of degrees of freedom for assessing the uncertainty of sample means. He proposed that the number of independent samples $N*$ contained in a time series of N time steps separated by Δt with an e-folding time of $T = -\Delta t / \ln(r\Delta t)$ is given by,

$$N^* = \frac{N\Delta t}{2T} = \frac{\text{Total time series length}}{\text{Two times the autocorrelation e-folding time}} \tag{7.14}$$

In other words, the number of degrees of freedom is the total length of the time sample divided by twice the e-folding time of the autocorrelation. Separation of two e-folding times between independent measurements is required since the intervening points can be mostly predicted by two points separated by a smaller time interval than $2T$. Leith's formula can also be written as,

$$\frac{N^*}{N} = \frac{1}{2}\ln(r(\Delta t)) \tag{7.15}$$

Leith's formula is consistent with Taylor (1921), who said that

$$\frac{N^*}{N} = \frac{1}{2L} \tag{7.16}$$

where L is given by

$$L = \int_0^\infty r(\tau')d\tau' \tag{7.17}$$

If we substitute the formula for the autocorrelation of red noise 7.8 into 7.17, and take into account that Taylor was using non-dimensional time $t' = t/\Delta t$, then we can show that $L = T$, and Leith's formula is the same as Taylor's.

The factor of two comes into the bottom of the above expressions for N* so that the intervening point is not easily predictable from the ones immediately before and after. If you divide the time series into units of e-folding time of the auto-correlation, T, One can show that, for a red noise process, the value at a midpoint, which is separated from its two adjacent points by the time period T, can be predicted from the two adjoining values with combined correlation coefficient of about $2e^{-1}$, or about 0.52, so about 25% of the variance can be explained at that point, and at all other intervening points more can be explained.

Bretherton et al. (1999) provide a nice review of efforts to assess spatial and temporal degrees of freedom. They use an approach in which the spatial or temporal statistics are fitted to a Chi-Squared distirbution. An alternative formula for effective degrees of freedom to be used in assessing the statistical significance of temporal means is given as,

$$\frac{N^*}{N} = \frac{(1 - r_1(\Delta t))}{(1 + r_1(\Delta t))} \tag{7.18}$$

If one is looking at a first order process, such as the calculation of a mean value, or the computation of a trend where the exact value of the time is know, then the formula 7.18 should be used.

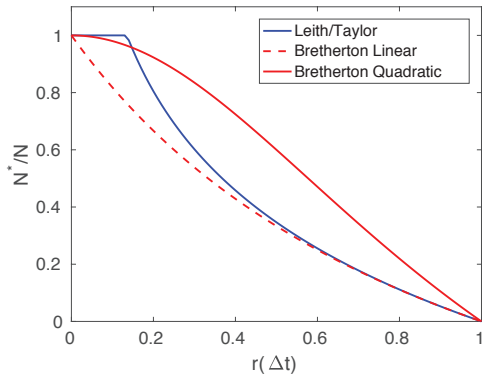

Figure 7.3 Ratio of degrees of freedom N^* to sample size N as a function of autocorrelation at one time step $r(\Delta t)$ for use in assessing the statistical significance of means(linear) and correlations (quadratic).

For quadratic statistics, such as variance and covariance analysis between two variables x_1 and x_2, a good approximation to use is:

$$\frac{N^*}{N} = \frac{(1 - r_1(\Delta t)r_2(\Delta t)}{(1 + r_1(\Delta t)r_2(\Delta t)} \tag{7.19}$$

where, of course, if we are covarying a variable with itself, then $r_1(\Delta t)r_2(\Delta t) = r(\Delta t)^2$. This formulation goes back as far as Bartlett (1935). Of course, if the time or space series is not Gaussian red noise, then the formula is not accurate. But it is still good practice to use it, and many geophysical time series contain a good measure of red noise.

Figure 7.3 shows $N*/N$ as a function of $r(\Delta t)$ for the formulas introduced in section 7.4. Because in quadratic statistics such as covariance, the noise is multiplied by itself, the reduction of degrees of freedom with increasing autocorrelation is slower than for linear statistics such as the mean value.

7.5 Harmonic Analysis and the Fourier Transform

Harmonic analysis is the interpretation of a time or space series as a summation of contributions from harmonic functions, each with a characteristic time or space scale. Consider that a variable $y(t)$ is defined on the interval $0 < t < T$. Then we can write any time series $y(t)$ as

$$y(t) = \sum_{k=0}^{\infty} \left[A_k \cos\left(2\pi k \frac{t}{T}\right) + B_k \sin\left(2\pi k \frac{t}{T}\right) \right] \tag{7.20}$$

where T is here the length of the period of record and k is a positive integer. $y(t)$ is a continuous function of t. The coefficients of this expansion in sines and cosines can be obtained by multiplying by a test function on the left, say $\cos\left(2\pi n \frac{t}{T}\right)$, where n is any positive integer.

Because the sines and cosines are orthogonal to each other on the interval $0 < t < T$, after we integrate over time we obtain simple algebraic equations for the coefficients.

$$A_0 = \overline{y} \qquad A_k = \frac{2}{T}\int_0^T y(t)\cos\left(2\pi k \frac{t}{T}\right)dt \qquad B_k = \frac{2}{T}\int_0^T y(t)\sin\left(2\pi k \frac{t}{T}\right)dt \qquad k > 0 \tag{7.21}$$

A time series of real data is usually presented at discrete values of time separated by a time step Δt. In that case the integrals presented in 7.21 are approximated with a summation over the time series consisting of a sample of N equally-spaced values.

$$A_0 = \overline{y} \qquad A_k = \frac{2}{N}\sum_{i=1}^{N} y(t_i)\cos\left(2\pi k \frac{i\Delta t}{T}\right) \qquad B_k = \frac{2}{N}\sum_{i=1}^{N} y(t_i)\sin\left(2\pi k \frac{i\Delta t}{T}\right) \qquad k > 0 \tag{7.22}$$

Here $T = N\Delta t$. Where the data are not equally spaced, A_k and B_k can be estimated by regression. In the case of equally spaced data, one can estimate N coefficients from N data points, but the $k = N/2$ is the maximum wavenumber that can be computed. $A_{k=0} = \overline{y}$ is the mean of the time series and $B_{k=0} = B_{k=N/2} = 0$. The highest wavenumber that can be computed, k=N/2, is for a wavelength of $2\Delta t$ for which the amplitude, but not the phase can be computed. The highest resolvable frequency $1/2\Delta t$, is half the sampling frequency and is called the "Nyquist" frequency after Harry Nyquist, a Swedish born American electronic engineer. The highest frequency resolved thus depends on the sampling interval. The lowest frequency resolved is $1/T$, and is thus determined by the total length of the sample. The separation between frequencies is also $1/T$, which can be called the "bandwidth" of the analysis. The frequencies resolved by a sample of N values separated by Δt are thus,

$$f_i = \frac{i}{T}; \qquad i = 0, 1, 2, ...N/2 \tag{7.23}$$

Since the sines and cosines are orthogonal on the interval $0 < t < T$, the time series can be reconstructed exactly from the Fourier coefficients.

$$y(t_i) = \overline{y} + \sum_{k=1}^{N/2} A_k \cos\left(2\pi k \frac{i\Delta t}{T}\right) + B_k \sin\left(2\pi k \frac{i\Delta t}{T}\right) \tag{7.24}$$

This can be rearranged slightly to be,

$$y(t) = \overline{y} + \sum_{k=1}^{N/2} C_k \cos\left(\frac{2\pi k}{T}(t - t_k)\right) \tag{7.25}$$

where,

$$C_k^2 = A_k^2 + B_k^2 \quad \text{and} \quad t_k = \frac{T}{2\pi k}\arctan\left(\frac{B_k}{A_k}\right) \tag{7.26}$$

In 7.26 the time series is represented by a summation of cosine waves, each with an amplitude C_k and a phase delay t_k.

7.6 The Power Spectrum

In many cases of interest, it is useful to know how the variance of a time series is distributed across the frequency domain. Having performed the Fourier analysis of a discrete time series and expressed it in terms of a set of cosine waves with different frequencies, we can easily express the variance of y in the following way.

$$\overline{y'^2} = \frac{1}{2}\sum_{k=1}^{N/2} C_k^2 \tag{7.27}$$

So the power as a function of frequency f_k can be written as,

$$\Phi(f_k) = \frac{1}{2}C_k^2 \qquad \text{where} \qquad f_k = \frac{k}{T} \tag{7.28}$$

From 7.27 and 7.28 we infer that if we plot the power spectrum as a function of frequency, then the area under the curve will be proportional to the variance, a useful thing. If the frequency range is very large, one can plot $f_k\Phi(f_k)$ versus $\log f_k$ and preserve the area - variance relationship. If the power range and frequency range are both very large, one can plot the log of power versus the log of frequency, but in that case the area under the curve is no longer proportional to variance, and one must be careful when interpreting the contribution of different frequencies to the total variance. Since the power spectrum is based on N data points, and the resulting power spectrum has only $N/2$ values, each realization of a power spectrum has about two degrees of freedom. We don't need to worry about autocorrelation, since spectral analysis takes this into account explicitly.

7.7 Methods of Computing Power Spectra

7.7.1 Direct Fourier Transform

The power spectrum can be obtained by direct Fourier transform, as described in section 7.5. This is feasible for large data sets because of the Fast Fourier Transform (FFT), which greatly speeds up the computation of Fourier transforms if the length of the record is a power of two, $N = 2^m$, where m is any integer. Mixed radix FFTs are also available for $N = 2^m 3^n 5^p$. As previously mentioned, however, each spectral estimate has only about 2 degrees of freedom. To give the estimate of the power spectrum more statistical robustness, one can average adjacent frequencies together, or divide the record up into shorter segments, and average the spectral estimates from these shorter segments into an average spectrum with more degrees of freedom. In either case, one is forced to consider a trade off between spectral resolution, which depends on the length of the chunk of data given to the spectral analysis, and the number of degrees of freedom in the final spectrum. If L spectral estimates result from the analysis, and N total data are used to make that estimate, then the number of degrees of freedom is slightly more than N/L. Generally, the more degrees of freedom in the final spectral estimate, the better the quality of the analysis. One should also insist on adequate quality.

7.7.2 Lag Correlation Method

Norbert Wiener showed that the autocovariance and the power spectrum are Fourier transforms of each other. So we can obtain the power spectrum by performing harmonic analysis on the autocovariance function. This

was the method preferred before fast computers, because the number of computations required is much less than a direct Fourier transform. The number of lags can be chosen to achieve the desired frequency resolution, and the number of degrees of freedom of the resulting power spectrum increases with the length of the available record. Suppose we consider time lags, τ on the interval $T_L < \tau < T_L$. The Fourier transform pair of the continuous spectrum and the continuous lag correlation are then,

$$\Phi(k) = \int_{-T_L}^{T_L} r(\tau) e^{-ik\tau} d\tau \tag{7.29}$$

$$r(\tau) = \int_{-k^*}^{k^*} \Phi(k) e^{ik\tau} dk \tag{7.30}$$

The maximum lag, T_L, which is $L\Delta t$ in discrete mathematics, determines the bandwidth of the spectrum and the number of degrees of freedom associated with each spectral estimate. The bandwidth is $1/2T_L$, and frequencies resolved are $k/(2L\delta t$, where $k = 0, 1, 2, ...L$. There are L spectral estimates produced, and if N data are used to compute them, then each spectral estimate has about N/L degrees of freedom. So if $N = 1000$ data points are used to compute spectral estimates at $L = 50$ frequencies, each estimate has 20 degrees of freedom.

The lag correlation method is rarely used nowadays, because Fast Fourier Transform algorithms are more efficient and widespread. The lag correlation method is important for intellectual and historical reasons, and because it comes up again in higher order spectral analysis.

7.8 The Complex Fourier Transform and Spectral Analysis

In previous sections we presented the Fourier Transform in real arithmetic using sine and cosine functions. It is much more compact and efficient to write the Fourier Transform and its associated manipulations in complex arithmetic. In a domain of continuous time and frequency, we can write the Fourier Transform Pair as integrals:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega) e^{i\omega t} d\omega \tag{7.31}$$

$$F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \tag{7.32}$$

Here $f(t)$ is some real time series in the independent variable t, and $F(\omega)$ is the Fourier Transform of $f(t)$, and is generally a complex number with a real and imaginary part. ω is the frequency in radians per unit time. If the period of the oscillation is T, then the radian frequency is $\omega = 2\pi/T$, and the frequency in cycles per unit time is $f = 1/T$.

7.8.1 Parseval's Theorem

The following theorem by Parseval is important in spectral analysis and filtering theory. It is derived by considering two functions $f_1(t)$ and $f_2(t)$ with Fourier Transforms $F_1(\omega)$ and $F_2(\omega)$. We consider the following manipulation.

$$\begin{aligned}\int_{-\infty}^{+\infty} f_1(t)f_2(t)dt &= \int_{-\infty}^{+\infty} f_1(t)\left[\frac{1}{2\pi}\int_{-\infty}^{+\infty} F_2(\omega)e^{i\omega t}d\omega\right]dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} F_2(\omega)\int_{-\infty}^{+\infty} f_1(t)e^{i\omega t}dtd\omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} F_2(\omega)F_1(-\omega)d\omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} F_1(\omega)F_2^*(\omega)d\omega\end{aligned} \tag{7.33}$$

Here the asterisk indicates a complex conjugate. Since the original time series are real, we must have that $F(-\omega) = F^*(\omega)$. In the special case where $f_1(t) = f_2(t) = f(t)$ we obtain,

$$\int_{-\infty}^{+\infty} f(t)^2 dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \left|F(\omega)\right|^2 d\omega \tag{7.34}$$

Thus the square of the time series integrated over time is equal to the square (inner product) of the Fourier transform integrated over frequency. This shows that the integrated variance in time is equal to the power spectrum integrated over frequency.

7.8.2 The Time Shifting Theorem

The time shifting theorem indicates that introducing a time shift in complex Fourier space is a simple multiplication by a complex number. To see this consider the Fourier transform of a time series shifted by a time increment τ.

$$f(t \pm \tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(\omega)e^{i\omega(t\pm\tau)}d\omega = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(\omega)e^{\pm i\omega\tau}e^{i\omega t}d\omega \tag{7.35}$$

From 7.35 we see that the Fourier transform of $f(t \pm \tau)$ is the Fourier transform of $f(t)$ multiplied by the factor $e^{\pm i\omega\tau}$.

7.8.3 Lagged Covariance and the Power Spectrum

The continuous form of the definition of the lag covariance function for a time series $f(t)$is,

$$r(\tau) = \int_{-\infty}^{+\infty} f(t)f(t+\tau)dt \tag{7.36}$$

We can use Parseval's theorem and the time shifting theorem to write,

$$\begin{aligned} r(\tau) &= \int_{-\infty}^{+\infty} f(t) f(t+\tau) dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega) F^*(\omega) e^{i\omega\tau} d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Phi(\omega) e^{i\omega\tau} d\omega \end{aligned} \tag{7.37}$$

where $\Phi(\omega)$ is the power spectrum. The power spectrum is thus the Fourier Transform of the autocovariance function, so that they form a Fourier transform pair.

$$\begin{aligned} r(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Phi(\omega) e^{i\omega\tau} d\omega \\ \Phi(\omega) &= \int_{-\infty}^{+\infty} r(\tau) e^{-i\omega\tau} d\tau \end{aligned} \tag{7.38}$$

This methodology can also be applied to the covariance between two different time series $f_1(t)$ and $f_1(t)$. A similar relationship occurs between the covariance between two time series and the cross power $\Phi_{12}(\omega)$. The cross power has a real part, or cospectrum, and an imaginary part, or quadrature spectrum, which together specify the phase between the two time series.

$$\begin{aligned} r_{12}(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Phi_{12}(\omega) e^{i\omega\tau} d\omega \\ \Phi_{12}(\omega) &= \int_{-\infty}^{+\infty} r_{12}(\tau) e^{-i\omega\tau} d\tau \end{aligned} \tag{7.39}$$

7.8.4 Example: The Power Spectrum of Red Noise

The autocorrelation function for red noise is,

$$r(\tau) = e^{-\tau/T} \tag{7.40}$$

where T is the e-folding time of the autocorrelation, *i.e.* autocorrelation time. Using 7.38 and inserting 7.40 we obtain,

$$\Phi(\omega) = \int_{-\infty}^{+\infty} e^{-\tau/T} e^{-i\omega\tau} d\tau = \frac{2T}{1+\omega^2 T^2} \tag{7.41}$$

Figure 7.4 shows that the power spectrum of red noise peaks strongly at low frequencies as the autocorrelation time, T, increases. Note that the range of frequency in this theory is infinite, so as $T \to 0$ $\Phi(\omega) \to 0$, but takes on a uniform value for $0 < \omega < \infty$. The total variance is unchanged. In Figure 7.4 we show only the range $0 < \omega < \pi$, which would be the Nyquist range for $\Delta t = 1$ using discrete data. Later we will consider how this theoretical spectrum for red noise is modified for a finite sample of discrete data.

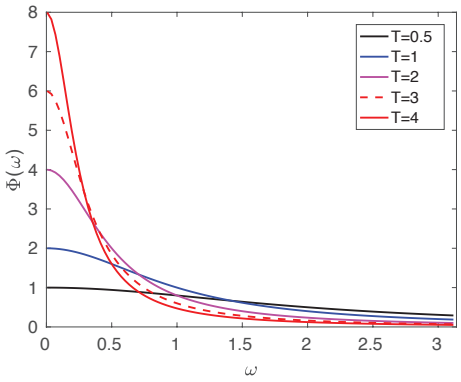

Figure 7.4 Power spectrum as a function of frequency in radians for the theoretical spectrum in 7.41 for autocorrelation e-folding times of 0.5, 1, 2, 3, and 4 time units.

7.9 Data Windows and Window Carpentry

In the analytic case we presume an infinite domain so that the true spectrum can be calculated exactly, provided the analytic function satisfies certain conditions.

$$F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \tag{7.42}$$

In real cases, however, where we cannot observe $f(t)$ on the interval $-\infty < t < \infty$, we are looking at the function through a "window." The window can be represented by a function $w(t)$. The window function for the ideal analytic case where we know the function for all time is $w(t) = 1$ on the interval $-\infty < t < \infty$, but in the more realistic case where we know the function only on some finite interval, say $-T/2 < t < T/2$, then $w(t) = 1$ on that interval and $w(t) = 0$ everywhere else. In the case of a finite window, we do not see the true time series $f(t)$, but that time series multiplied by a window function $f(t)w(t)$. In order to understand the effect on the Fourier transform of multiplying the time series by a window function, we can use a form of the convolution theorem that states that the Fourier transform of the product of two functions is the convolution of their individual Fourier transforms 7.43.

$$\int_{-\infty}^{+\infty} f(t) w(t) e^{-i\omega t} dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\overline{\omega}) W(\omega - \overline{\omega}) d\overline{\omega} \tag{7.43}$$

It is useful then to know the Fourier transform of the window function to see how it modifies the true Fourier transform, which is related closely to the power spectrum by 7.34. Consider the following Boxcar function window.

$$w(t) = \begin{cases} 1/T & \text{for} \quad T/2 < t < T/2 \\ 0 & \text{otherwise} \end{cases} \tag{7.44}$$

The Fourier transform of the boxcar function 7.44 is easily obtained.

$$\begin{aligned} W(\omega) &= \int_{-\infty}^{+\infty} w(t) e^{-i\omega t} dt = \frac{1}{T} \int_{-T/2}^{T/2} e^{-i\omega t} dt \\ &= \frac{1}{i\omega T} \left[e^{i\omega T/2} - e^{-i\omega T/2} \right] = \frac{\sin\left(\frac{\omega T}{2}\right)}{\frac{\omega T}{2}} = \text{sinc}\left(\frac{\omega T}{2}\right) \end{aligned} \tag{7.45}$$

The Fourier transform of the boxcar function is a sinc function, $\mathrm{sinc}(x) = \sin(x)/x$, which has the unfortunate characteristics of relatively large negative side lobes that decay slowly, so that it spreads variance around in frequency space in a spurious way. Figure 7.5 shows this behavior. The first zero crossing of the response function $W(\omega)$ occurs at $\omega = 2\pi/T$, which for a discrete Fourier analysis is the lowest frequency resolved and also the separation between resolved frequencies (*i.e.* the bandwidth). The first negative side lobe always has an amplitude of about -0.22, so a spurious signal of about 22% appears a bandwidth or so away from the actual frequency.

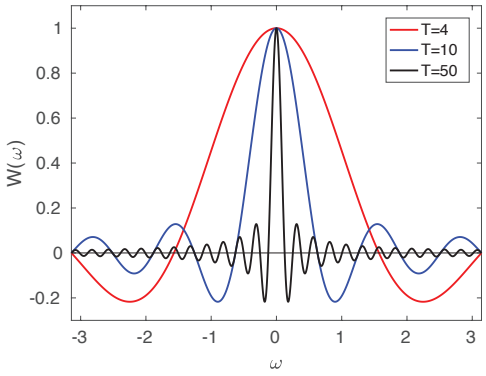

Figure 7.5 Fourier transform of the rectangular window function for various values of the length of record T in time units. The plot is terminated at $\omega = \pm\pi$, which would be the Nyquist interval in radian frequency if the time step was one time unit.

A finite window acts as a smoothing on the true spectrum. In addition to the smoothing effect, the side lobes of the frequency window lead to spectral leakage to other frequencies. The degree of smoothing and the range of the spread in frequency depend on the length of the data window T. The shorter that T is, the stronger the smoothing/spread will be. Therefore we can see that we must carefully design the window through which we view the data prior to spectral analysis, if we want to obtain the best results. The response function for the rectangular window is shown below. Note that while the response does peak strongly at the central frequency, significant negative side lobes are present. This means that our spectral analysis will introduce spurious oscillations at higher and lower frequencies that are out of phase with the actual oscillation.

To improve the quality of the resulting spectral analysis, it is desirable to modify the data window from the naive rectangular window. The ideal window response function, $W(\omega)$, would have a narrow central lobe and insignificant side lobes. We can improve on the naive rectangular window function and the rather unsatisfactory sinc function window response through modifications of the window function $w(t)$. Since much of the poor behavior of the rectangular window comes from the Gibbs effect of its sharp edges, rounding the edges of the window is an intuitive approach. Many data windows have been proposed for different purposes, most of which taper the window to near zero at the ends and maximize in the middle of the data window Harris (1978). In addition to the rectangular window already discussed, we will introduce just two that are suitable for general use; the Hann or Hanning window, and the Hamming window.

7.9.1 The Hanning Window

The Hanning window is named after Julius von Hann and is also called the elevated cosine or cosine bell window.

$$w(t) = \frac{1}{2}\left(1 - \cos\left(\frac{2\pi t}{T}\right)\right) \quad T/2 < t < T/2 \tag{7.46}$$

$$W(\omega) = \mathrm{sinc}\left(\frac{\omega T}{2}\right) + \frac{1}{2}\left(\mathrm{sinc}\left(\frac{\omega T}{2} + \pi\right) + \mathrm{sinc}\left(\frac{\omega T}{2} - \pi\right)\right) \tag{7.47}$$

We can see that the first part of the response function for the cosine bell window is a sinc function exactly like that of the rectangular window. In addition, however, we have two additional sinc functions of reduced amplitude that maximize in the center of the negative lobes on either side of the central lobe. The effect of these is to nearly cancel the negative side lobes of the rectangular response function while significantly broadening the central lobe (Fig. 7.6). We like the fact that the side lobes are now smaller, but the broadened central lobe means that the spectrum will be slightly smoothed compared to a rectangular window. This smoothing is not a disaster in most applications, since we often end up doing some smoothing anyway, and the smoothing effect of the cosine bell window allows us to claim a small increase in the number of degrees of freedom per spectral estimate. If greater frequency resolution is required, then a longer chunk of data must be used (*i.e.* increase T)

7.9.2 The Hamming Window

If a slightly better cancellation of the side lobes is desired, then the Hamming window (Richard W. Hamming) can be used. In addition to slightly better cancelling of the side lobes, a narrower central lobe is achieved compared to the Hanning window. Figure 7.6 shows the Fourier transforms of the Rectangular, Hanning and Hamming windows for a window length of $T = 100$. The side lobes are greatly reduced by tapering the window. The differences between the Hanning and Hamming windows are modest.

$$w(t) = \frac{1}{2} - 0.426\cos\left(\frac{2\pi t}{T}\right) \quad T/2 < t < T/2 \tag{7.48}$$

$$W(\omega) = \mathrm{sinc}\left(\frac{\omega T}{2}\right) + 0.426\left(\mathrm{sinc}\left(\frac{\omega T}{2} + \pi\right) + \mathrm{sinc}\left(\frac{\omega T}{2} - \pi\right)\right) \tag{7.49}$$

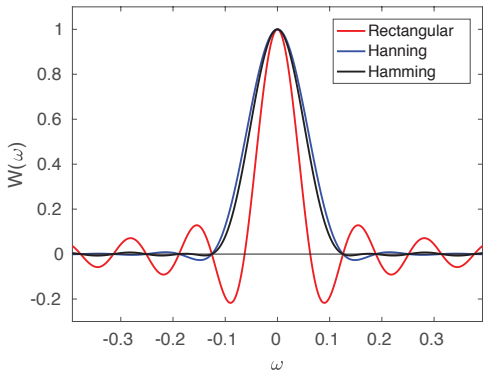

Figure 7.6 Response functions $W(\omega)$ for the rectangular, Hanning and Hamming windows for a window length of $T = 100$.

7.9.3 Welches Overlapping Segment Analysis: WOSA

The most common methods of spectral analysis used employ the Fast Fourier Transform method. In this method a direct Fourier Transform is made of the data using an efficient algorithm that makes use of the fact that the length of the time series has been chosen to be an integer power of two $M = 2^n$. Mixed-

radix FFT's are also available for which $M = 2^n 3^m 5^j$. In applying these methods the total time series of length $N\Delta t$ can be broken up into a series of smaller chunks of length M. Since a tapered window like the Hanning window will normally be applied, it is better to overlap the segments so that the data near the break points are not "wasted" by receiving a small weight. Overlapping the data by 50% will ensure that all the data are counted equally in the average spectrum that will be accumulated by averaging the results from each individual segment. This averaged spectrum will have approximately 2N/M degrees of freedom, since each power spectrum will have only M/2 estimates. The number of degrees of freedom will actually be slightly larger than this, depending on how much smoothing the data window provides, making the number of independent spectral estimates for each realization of the spectrum smaller than M/2. The spectra and cross-spectra for these smaller chunks can be averaged into a grand spectrum that has some degree of statistical reliability if N >> M. This is called Welch's Overlapping Segment Analysis, or WOSA.

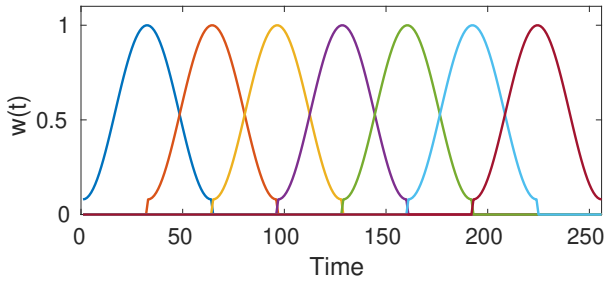

Figure 7.7 Illustration of WOSA analysis in which Hamming windows of length M = 64 are overlapped by 50% on a data set of length N = 256.

Figure 7.7 illustrates how a set of Hamming windows of length M = 64 overlapped by 50% can be used to analyze a data set of length N = 256 . All the data are equally weighted in the composite spectrum that results, except for the data on either end of the data set that are under weighted. Since the resulting spectrum will contain only M/2 spectral estimates, and N data are used, the number of degrees of freedom per spectral estimate is approximately 2N/M, times some factor to take into account that the spectrum is smoothed by the window. In the case of the Hamming window this factor is about 1.2.

Figure 7.8 shows example results for computing a spectrum using WOSA analysis and three different windows, the rectangular window, the Hanning taper and the Hamming taper. The input time series is a cosine wave with a period of 4.2 time units. The total record is of length N = 2048, a chunk length of M = 64 was used with an overlap of 50% = 32 time units. The power scale is logarithmic to show the sensitivity where the estimated power is very small. As expected from 7.6 one can see that the tapered windows greatly reduce the amplitude away from the line center at $\omega = 2|\mathrm{pi}/2.4$, especiallly for the first side lobe, but they also widen, smooth out, the central peak to be twice qs wide as for the rectangular window. The Hamming window does a more effective job than Hanning of removing the first side lobes, which are the biggest, but do allow more variance to pass far away from the line center. These amplitudes are weaker by a factor of 10^{-4} from the peak power and would not be a concern in typical geophysical settings with lots of noise.

7.10 Designing a Power Spectral Analysis

When considering spectral analysis of a time or space series, one typically has a hypothesis that a peak in the spectra my occur in a particular frequency range, which would indicate a larger than expected amount of variance with the corresponding period. It is essential to establish an *a priori* argument for where that spectral peak should be. Once this is known the nature of the required data set and an effective approach to spectral analysis can be designed.

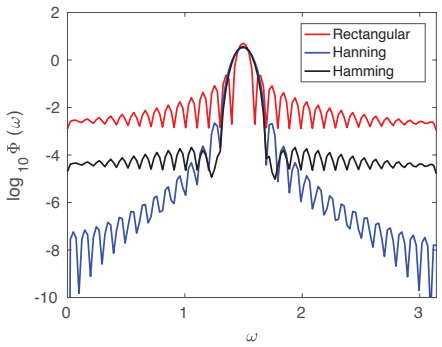

Figure 7.8 Power spectrum of a cosine wave with a period of 4.2 time units estimated with WOSA analysis in which Hamming windows of length $M = 64$ are overlapped by 50% on a data set of length $N = 2048$. The results with a rectangular window, a Hanning taper and a Hamming taper are shown.

7.10.1 Bandwith and Chunk Length

The first thing that to consider is how much resolution of frequency is required to isolate the peak of interest. Is it a very sharp peak, at a particular frequency, or is it expected to be a broad peak? If another peak is expected to be nearby, how much bandwidth do you need to separate the peak you seek from others you know to be around. The frequency separation between a period of P and a period of P' is $\Delta f = 1/P' - 1/P$. Supposing that you want at least a couple frequencies in between to show the separation between these peaks, and you are using a finite window that will smooth the spectrum a bit, you probably want a bandwith of at most one fourth of the frequency separation between the periods of interest. The bandwith of a spectral analysis is $\Delta f = 1/M\delta t$, so you want a chunk length of at least four times the longer of the two periods P and P'.

7.10.2 Time Step

Suppose we know that we expect a peak in the variance at a period of P time units, or a frequency of $f = 1/P$ cycles per unit time. A time step of $\Delta t = P/4$ will put that spectral peak right in the center of the Nyquist interval $0 < f < 1/2\delta t = 2/P$. So half the resolved frequencies will be higher than the frequency of interest and half will be lower. There is little point in having a smaller time step than this, since it just adds more high frequencies and does not help at all with the frequency of interest. Being in the middle of the Nyquist interval is more than enough to reduce any problems with aliasing from frequencies higher than those resolvable by the time step chosen.

7.10.3 Robustness and Degrees of Freedom

In designing a spectral analysis procedure, one must take into account the tradeoff between spectral resolution and degrees of freedom, if the length of the available time series is limited. We stated before that the number of degrees of freedom in a spectral estimate is approximately the number of data points divided by the number of independent spectral estimates. The number of degrees of freedom required depends on the relative strength of the peak we are looking for. We will come back to this in the section on testing the statistical significance of spectral peaks, but for now let's assume a general rule that we don't take a spectrum seriously unless we have about 20 degrees of freedom, so this means we need a data set that is 10 times the length of the chunk, $N = 10xM$. With that let's consider an example to fix ideas.

7.10.4 Example of 5-Day Wave

Suppose we have an *a priori* expectation that a peak in variance will occur at 5 days. What is the time step Δt and chunk length M that will resolve this well, and how much data do we need to get a robust result? Our frequency of interest is $f = 0.2$ cycles per day (cpd), and a time step of one day will give a Nyquist frequency of $f_{Nyquist} = 0.5$ cpd. Since it should be easy to get daily data, let's choose that as our sampling interval.

Next we need to decide what bandwith to use. Suppose for the sake of argument that we want to be able to distinguish a 5-day wave from a six day wave. So, $\Delta f = 1/5 - 1/6 = 1/30$. If we want three frequencies in between, then we need a bandwidth of about $\Delta f == 1/30x4 = 1/120$. Since we'll be using an FFT and WOSA analysis, let's choose a chunk length of $M = 128$, which is the nearest power of 2 greater than 120. To get about 20 degrees of freedom, we'll need 1280 days of data, or about 3.5 years.

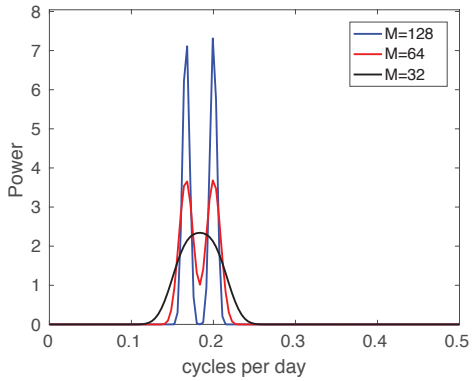

Figure 7.9 Power spectrum of a data set consisting of a 5-day and 6-day harmonic, using a Hamming taper with chunk length of M =128, 64 and 32. In this plot the total variance is the area under the curve, which is the same in each case.

Figure 7.9 shows the spectra computed from an input time series consisting of cosine waves of 5 and 6 days using our design chunk length of M =128, as well as shorter chunks of 64 and 32. As we expected, the 128-day chunk length allows the 5 and 6-day waves to be separated, with zero variance showing at 2 intervening frequencies. The separation is still observable with a chunk length of 64, but at a chunk length of 32 the two peaks merge into one broad peak.

7.11 Statistical Significance of Spectral Peaks

The statistical significance of a peak in a power spectrum is assessed as in any case by stating the significance level desired, and then stating the null hypothesis and its alternative. The null hypothesis is usually that the time series is not periodic in the region of interest, but simply noise. Since most geophysical time series contain a good measure of noise, we can usefully compare the amplitude of a spectral peak to a background value determined by a red noise fit to the spectrum. We illustrate here one simple method of determining if a spectral peak is statistically significant.

One way to test significance of a spectral peak is to compute the ratio of the observed power, Φ to the power expected from your null hypothesis Φ_0 and compare this value with a "Chi-Squared" test with the corresponding number of degrees of freedom,

$$\chi^2 = (n-1)\frac{s^2}{\sigma^2} \quad \nu = (n-1) \tag{7.50}$$

where the Number of degrees of freedom ν is one less than the number of independent samples, n. The number of independent samples is estimated from,

$$n = \frac{2N}{M} f_w \tag{7.51}$$

where,

$$\begin{aligned} N &= \text{total sample size} \\ M &= \text{chunk length of FFT} \\ f_w &= \text{a factor to account for the smoothing by the window} \end{aligned} \tag{7.52}$$

The factor of two arises because the spectrum has only $M/2$ estimates, since the phase information is not included in the power spectrum. The factor f_w is generally between 1 and 1.5, depending on how much smoothing of the spectrum that the particular window does. For the Hanning window, $f_w = 1.2$, To be on the conservative side, we can assume $f_w = 1.0$, but in marginal cases this factor can be used to get a more accurate significance estimate.

A more convenient way to apply this test is to use the F-test, based on the F distribution.
Theorem: If σ_1^2 and σ_2^2 are the variances of independent random samples of size n_1, and n_2, respectively, taken from two normal populations having the same variance, then

$$F_{\nu_1}^{\nu_2} = \frac{\sigma_2^2}{\sigma_1^2} = \frac{\Phi}{\Phi_0} \tag{7.53}$$

is a value of a random variable having the F distribution with the parameters $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$, giving the number of degrees of freedom for the upper and lower variance, respectively.

The F distribution can be used to determine whether two sample variances are different in a statistical sense at a chosen probability level. Tables of the F distribution are included in **??**. You should not confuse this F with $F(z)$, the cumulative distribution of the Normal distribution. In assigning a confidence level and interpreting statistical tests of significance one must be concerned very much with the distinction between *a priori* and *a posteriori* statistics.

a priori: If we have stated in advance that we expect a peak at a particular frequency (and given a good reason beforehand), then we can simply test the significance of the spectral peak above the background using the normal confidence limits set forth for the chi-squared or F statistics.

a posteriori: If we have not stated at which frequency we expect the peak, then we must determine the probability that one frequency out of the $M/2$ we have computed should show a significant peak. The usual way to do this would be to take the probability of a type II error (accepting a false hypothesis as true) and multiply this by the number of chances we have given the spectrum to exceed the required level.

7.11.1 Example: a priori versus a posteriori Spectral Peaks

Suppose we have a spectral peak which exceeds the background significantly at the 95% probability level.

1. If we had predicted this frequency before computing the spectrum, then we can use the 95% probability level and infer that only a 5% possibility exists that this spectral peak could have occurred by chance. We have an *a priori* reason for expecting this peak and we can therefore use *a priori* statistics.

2. If we had not predicted the frequency of the peak, then we must test the probability that one frequency out of our sample of $M/2f_w$ independent estimates should show a significant peak. $M/2$ is the number of frequencies retained in our spectrum and f_w is the factor indicating the degree of smoothing by the window. If the number of independent frequencies in our spectrum is 50, then our chance of getting a spectrum with no significant peaks is $(0.95)^5 0 = 0.08$ and it is likely that at least one peak will exceed the 95% confidence level by chance. If we had started with a 99% confidence limit the significance would be $(0.99)50 = 0.6$, and there is still a 40% chance that one peak will exceed the 99% confidence level. Therefore, in practice we need to have an *a priori* reason for expecting a peak at the frequency where one occurs, or our confidence

level is very low. This *a priori* reason could be a theory, or the observation of a peak at the frequency of interests in an independent data set. The probability of getting spectral peaks with 95% significance at the same frequency by chance in two independent data sets is small.

7.11.2 Example: Statistical Design

Example: Suppose we wish to examine climatic fluctuations in the 10-1000 year range. Assume that a physically meaningful peak should have twice the variance of the background spectrum. What should be the spacing of the observations and how long a time series is required to get meaningful statistics?

A time step of two years would resolve the shortest period of interest very well, since the Nyquist period would be 4 years and a period of 8 years would be in the middle of the Nyquist interval ($0 < f < 1/(2\Delta t)$. The next question would be how much bandwidth is required, which determines the chunk length, M. Suppose for the sake of argument that we'd like to be able to distinguish a 500 year period from a 1000 year period. If we choose a chunk length of 4096 years, then the fourth frequency resoloved would be 1/1024 cycles per year, and the eighth frequency would be 1/512 cycles per year. This is fine because it leaves three frequencies lower than the lowest frequency of interest, and it leaves three frequencies between the two frequencies of interest 1/1000 and 1/500 cycles per year. Since the time step is two years, we need a chunk length of $M = 2048$. To see how many degrees of freedom we need, we first need to decide what our significance level is. Let's say we want 99% significance if our spectral peak exceeds the background by a factor of two. We then ask how many degrees of freedom are required so that an F-statistic of 2.0 is 99% significant. We need to know the number of degrees of freedom for our null hypothesis spectrum. This is usually a two-parameter fit, so let us assume we have at least 100 degrees of freedom for it. Then from the F-statistic table, we determine that the sample spectrum must have at least 23 degrees of freedom. This means we need about 12 independent realizations of our spectrum or about 12x2x2048 = 49,152 years of data. So if we find a spectral peak with a variance ratio of 2 using a chunk length of 4096 years, then this peak will be significant at 99%, if we've predicted the frequency of the peak *a priori*. Two-year averages or data taken every other year are fine for this analysis. Using yearly data would just double the Nyquist frequency and add many high frequencies of no interest to our spectrum.

7.11.3 The Red Noise Null Hypothesis

A useful null hypothesis for many geophysical time series is that the time series consists of autocorrelated Gaussian noise. The degree of autocorrelation can be a very important physical characteristic and the reasons why geophysical time series are autocorrelated are interesting, but here we focus on the exceptions where the time series contains some periodic phenomena immersed in autocorrelated noise. The theoretical spectrum for autocorrelated "red" noise was presented in 7.2.3 , but here we consider how the red noise spectrum is modified by being viewed at discrete times separated by Δt. We begin with the equation for an autocorrelated random walk 7.3.

$$x(t) = \alpha x(t - \Delta t) + \left(1 - \alpha^2\right)^{1/2} \epsilon(t) \tag{7.54}$$

where $\alpha = r(\Delta t)$ Using the "time shifting theorem" 7.35, we can write the Fourier transform of 7.54 as,

$$\begin{aligned} X(\omega) &= \alpha X(\omega) e^{-i\omega\Delta t} + \left(1 - \alpha^2\right)^{1/2} E(\omega) \\ &= \frac{C E(\omega)}{1 - \alpha e^{-i\omega\Delta t}} \end{aligned} \tag{7.55}$$

where $X(\omega)$ is the Fourier transform of $x(t)$, $E(\omega)$ is the Fourier transform of Gaussian white noise, and $C = \left(1 - \alpha^2\right)^{1/2}$. We know from Parseval's Theorem that the power spectrum of $x(t)$ is

$$\Phi_{xx}(\omega) = X(\omega)X^*(\omega) = \frac{C^2 E(\omega)E^*(\omega)}{\left(1-\alpha e^{-i\omega\Delta t}\right)\left(1-\alpha e^{i\omega\Delta t}\right)} = \frac{1-\alpha^2}{1-2\alpha\cos(\omega\Delta t)+\alpha^2} \quad \text{for} \quad 0<\omega<\frac{\pi}{\Delta t} \tag{7.56}$$

where we have used the identity $\cos(x) = \frac{\left(e^{ix}+e^{-ix}\right)}{2}$. The formula 7.56 for the power spectrum of red noise was discussed by Gilman et al. (1963).

To fit the shape function 7.56 to a real spectrum we need the one-lag autocorrelation, α from the input time series. A slightly more robust estimate of the parameter α from the original time series is the average of the one-lag autocorrelation and the square root of the two-lag autocorrelation. It is presumed here that the time step is chosen appropriately for the variability present in the time series. If the time step is too small or too large for the characteristic variability in the time series, then the results will be poor. We then multiply the shape 7.56 for that value of α by a factor that will make the variance equal to that of the time series in question. One simple way to do this is to match the total variance. The total variance is the sum of the power over all non-zero frequencies. So sum the power in the observed spectrum and the power in the idealized red noise spectrum. Then multiply the idealized red noise spectrum by this ratio, so that the red noise spectrum has the same total variance as the observed spectrum. The null-hypothesis spectrum is thus a two-parameter fit that has the same total variance and same one-lag autocorrelation as the observed time series.

7.11.4 Continuous and Discrete Red Noise Spectra

It may be helpful to compare the discrete red noise spectrum 7.56 with the continuous red noise spectrum in section 7.8.4 . For the continuous spectrum 7.4 the integration is performed for $0 < \omega < \infty$, whereas for the discrete spectrum 7.56 the integration is performed for $0 < \omega < \pi/\Delta t$. The discrete spectrum has an integral of π, so that if $\alpha = 0$ the spectrum has a uniform value of 1. For the continuous spectrum, $\alpha = 0$ gives a vanishingly small value that stretches to infinity, although the total variance is conserved as α increases and the variance peaks near $\omega = 0$. To compare the shapes of the two formulas in the Nyquist interval, we can integrate the continuous spectrum over the Nyquist interval to establish its norm, and use that to rescale the continuous spectrum so that it has the same variance as the discrete spectrum on the Nyquist interval.

$$\int_0^{\pi/\Delta t} \frac{2T}{(1+\omega^2 T^2)} d\omega = -2\tan^{-1}\left(\frac{\pi}{\ln\alpha}\right) \tag{7.57}$$

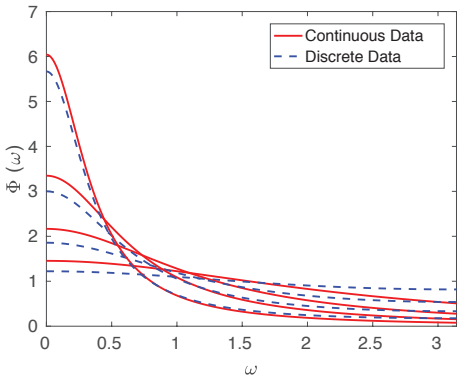

Figure 7.10 Power spectra of red noise computed using the continuous 7.4 and discrete 7.56 formulations. Curves are shown for $\alpha = 0.1$, 0.3 0.5 and 0.7. The continuous spectrum is normalized for the Nyquist interval using 7.57.

Comparison of the continuous and discrete approximations of the power spectrum of red noise are shown in Figure 7.10. It can be seen that the discrete approximation smooths the spectrum a little, with lower values at low frequencies nd higher values at high frequencies. The discrete approximation will be used to test experimental spectra computed with discrete data.

7.11.5 Example: Sampling Red Noise

In this section we consider samples of red noise and use the red noise null hypothesis and F-statistic to test for the significance of the peaks that appear. This example shows that one sample of red noise can show a spectrum that appears to have peaks, but proper use of statistics shows these peaks to be insignificant. As an example, we generate a time series of daily observations with an autocorrelation of $\alpha = 0.5$ at a one-day lag. We choose a chunk length of 256 days, which yields a spectrum with 128 estimates. To assess statistical significance we use a red noise spectrum with the shape of 7.56 and the same variance as the observed spectrum. We assess the number of degrees of freedom in the spectrum and then multiply the null-hypothesis spectrum by the appropriate F-statistic.

$$\Phi_{99} = \Phi_{\mathrm{Null}} \times \mathrm{F}_{\nu_0}^{\nu 0.01} \tag{7.58}$$

Here the F-statistic has three parameters, the number of degrees of freedom in the null hypothesis ν_0, the number of degrees of freedom in the sample ν, and the significance level, $p = 0.01$ for 99% confidence. To increase the degrees of freedom and improve the robustness of the spectrum we can average multiple realizations of the spectrum by using independent chunks of data.

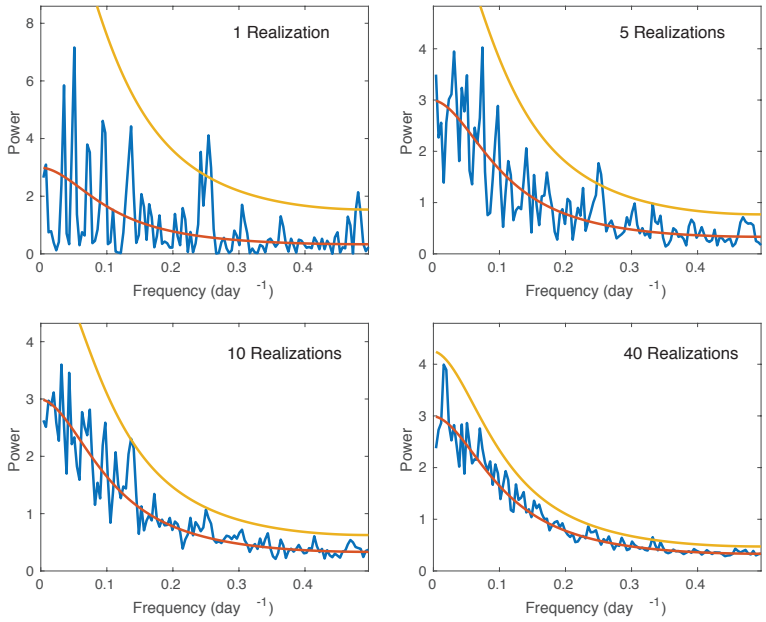

Figure 7.11 Power spectra of red noise with an autocorrelation of $\alpha = 0.5$ at one day, using a Hamming window and a chunk length of M=256 days. Examples are show for averages over 1, 5, 10 and 40 realizations. The red line is best fit using 7.56. The orange line is the 99% confidence limit, calculated as explained in the text.

Figure 7.11 illustrates the impact of averaging multiple realizations of a spectrum. A single realization shows a peak around f =0.25 cycles per day that exceeds the 99% confidence level indicated by the orange line. Since the spectrum has 128 chances to exceed the 99% confidence level, it is not surprising that one or two frequencies exceed the 99% threshold. Since we did not predict a peak at f =0.25 and we know that the observed spectrum has only about 2 degrees of freedom, we would not take this peak seriously. It is highly

likely that a sample of red noise will produce a spikey spectrum by chance. Since it got off to such a good start, by chance, with the first realization, the spectral peak around $f = 0.25$ still passes the 99% threshold of significance when 4 more realizations are added. It is still not significant, since we have no *a priori* reason to expect it in the spectrum with 5 realizations. When we add 5 more realizations for 10 total, the $f = 0.25$ peak finally is below the significance level, but a new peak at $f = 0.13$ appears to touch the 99% threshold, and again we have not predicted that peak in advance, so we would not take it seriously without further analysis. If we include a total of 40 realizations the two peaks previously considered fall below the line, and another one jumps out around $f = 0.34$. Again, we expect 1 or 2 frequencies out of 128 to pass the 99% threshold by chance. An extremely large sample would be needed to get all of the kinks and wiggles out of the experimental spectrum. The lesson here is to insist on sufficient quality and don't become excited by spurious peaks that appear by chance.

7.12 Prewhitening

The previous section discussed one method for determining the significance of spectral peaks emersed in red noise. It is also possible to remove the red noise from a data set prior to spectral analysis, a process called "prewhitening" the time series. The approach is very simple, one just subtracts the red noise approximation to the time series. A prewhitened time series $x_w(t)$ is computed from $x(t)$ in the following way.

$$x_w(t) = x(t) - \alpha x(t - \Delta t) \tag{7.59}$$

Here $\alpha = r(\Delta t)$. To illustrate how prewhitening can work we consider red noise with an autocorrelation of $\alpha = 0.5$ for daily sampling, and we add pure periodicities with periods of 10 and 4 days. Figure 7.12 considers red noise as in Figure 7.11, but adds periodicities at 4 and 10 days. To illustrate the effect of chunk length, we include spectra computed with $M = 256$ and $M = 128$. To be fair, we give the shorter chunk length twice as many realizations, so it has the same amount of data to work with. In this case with only red noise and periodicities in the time series, prewhitening works beautifully and produces a white noise spectrum with periodicities. The significance assessment is the same, however, and both methods and both chunk lengths properly identify the periodicities as significant, assuming we had an *a priori* reason for expecting peaks at 4 and 10 days. Note that the noise used in each of the cases in Figure 7.12 was the same realization, but different from the noise realization used in Figure 7.11.

7.12.1 Rossby-Gravity Waves in Reanalysis

Spectral analysis is an effective tool for identifying phenomena with a well-defined period. In the days before global weather analyses, spectral analysis was used to identify waves in time series from a few stations where balloon observations of wind and temperature were available. A classic example is the definition of tropical waves from rawinsonde stations in the tropical Pacific. Examples include Kelvin waves (Wallace and Kousky, 1968), Mixed Rossby-Gravity waves (Yanai et al., 1968) and the Madden-Julian Oscillation (Madden and Julian, 1971). Nowadays reanalysis products reconstruct global maps of wind, temperature and other meteorological variables, even where few observations are present, by simulating the dynamics and thermodynamics of the atmosphere as part of the interpolation process. Here we will look for the Mixed Rossby-Gravity wave discovered theoretically by Matsuno (1966) and later observed by Yanai et al. (1968). We expect the MRG wave to have a periodicity around 4-5 days in the meridional (north-south) wind at the equator. We will look for it in the central equatorial Pacific between 190E and 210E. We use data every 5-degrees of longitude and average the spectra together. Data for calendar years 2000-2015 are used and the annual cycle is removed.

Figure 7.13 shows power spectra of the meridional wind at 850hPa at the equator using three different methodologies, plus a contour plot of the fraction of variance in the 3 to 6-day period band. Our expectation is that a spectral peak will appear near 4 to 5 days associated with the theoretical prediction of the Mixed

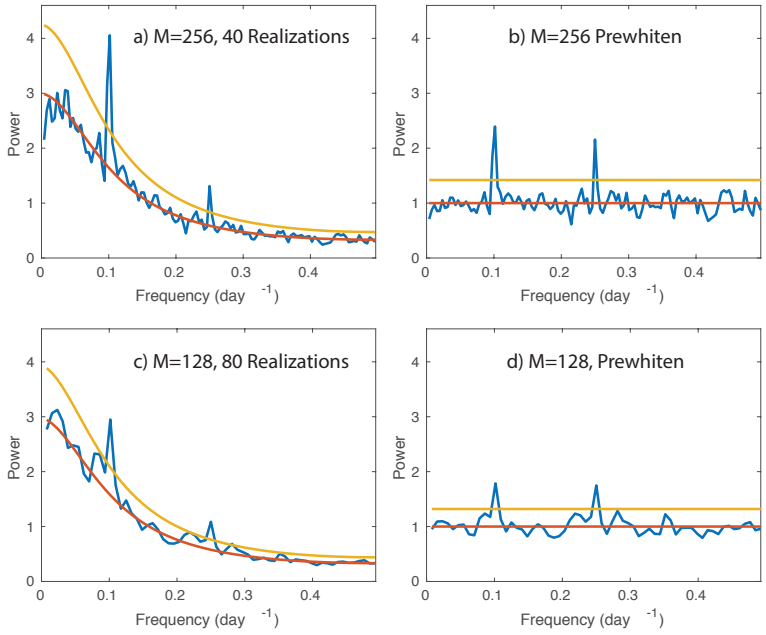

Figure 7.12 Power spectra of red noise with an autocorrelation of $\alpha = 0.5$ at one day plus harmonics with periods of 4 and 10 days. Spectra are computed using a Hamming window and chunk lengths of M=256 and M=128 days. Spectra with and without prewhitening are shown. To be fair, the same number of data are presented to each analysis, so the shorter chunk length of 128 days is given 80 realizations compared to 40 for the longer chunk length. The orange line is the 99% confidence limit, calculated as explained in the text.

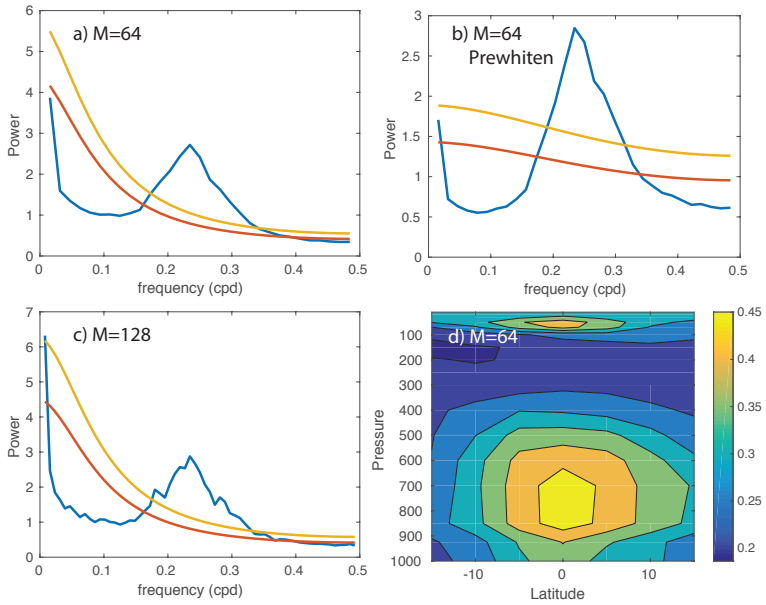

Figure 7.13 a) Power spectrum of the meridional component of wind at the equator and 850hPa averaged over the longitudes of 190-210E using a chunklength of a) $M = 64$, b) M=64 with prewhitening, c) $M = 128$. d) contour plot of the fraction of the total power that is contained in the 3-6 day period band as a function of latitude and pressure, computed with $M = 64$ without prewhitening. Red and orange curves indicate the red noise null hypothesis and 99% confidence level.

Rossby-Gravity wave. To distinguish 4 from 5 days we need a bandwith of $\Delta f = \left(\frac{1}{4} - \frac{1}{5}\right)/3 = \frac{1}{60}$, which will be nicely accomplished with a chunk length of $M = 64$. For comparison, we also show a chunk length of $M = 128$, which shows some additional detail that does not appear to be significant. Because the timeseries has such a strong peak in variance near 4.5 days, a red noise fit is not a good fit to the spectrum. If we remove our best fit to red noise from the time series by prewhitening, then we get a slightly more pleasing appearance (Fig. 7.13b), but no real difference in interpretation. Some autocorrelation remains after the red noise is removed because the periodicity itself will contribute autocorrelation . Plotting the ratio of the variance in the 3 to 6-day period range to the total variance (Fig. 7.13d) shows that the periodicity is confined near the equator and to the lower troposphere, although it does appear again in the lower stratosphere.

7.13 Multi-Taper Method of Spectral Analysis

In the Welch's Overlapped Segment Averaging (WOSA) method described in section 7.9.3, a single window function is applied to different segments of the time series and then averaged to produce a final spectral estimate with robustness and good spectral response properties. The choice of chunk length and window shape is based on the spectral resolution and degrees of freedom desired. Different chunk lengths and window shapes give optimal performance for different ranges of frequency. As the name implies, the multi-taper method (Thomson, 1982) uses a set of different data tapers to try to provide an optimal estimate of the spectrum of the time series. A nice description is given in Percival and Walden (1993). The idea is to choose a set of orthogonal tapers that provide optimal resolution and minimum leakage. A set of such tapers is devised that are harmonics on the data interval, and then the average of spectra computed using a number of these tapers is formed. Such a technique can give better results in some cases.

7.14 Maximum Entropy Spectral Analysis

Maximum entropy spectral analysis is a technique that can be used when you have a short period of record, but you need more spectral resolution that you can get by doing traditional Fourier Spectral analysis on the available data (Marple, 1987). It provides this extra spectral resolution by extending the autocovariance matrix in a way that adds the least information to the covariance matrix (maximum entropy). It will tend to strongly localize spectral peaks, so you can determine their location very precisely. The problem is that it has adjustable parameters that can be used to get a spectrum of arbitrary sharpness, and it may split peaks if the order of approximation is too high. The tools for assigning statistical significance to the results of such an analysis are uncertain. It is best used in conjunction with traditional Fourier spectral analysis, after you have established the significance of periodicities in the time series of interest.

7.15 Cross Spectrum Analysis

Cross spectral analysis allows one to determine the relationship between two time series as a function of frequency. Normally, cross-spectral analysis makes sense when statistically significant peaks at the same frequency have been shown in two time series. In that case we wish to know if these periodicities are related with each other and, if so, what the phase relationship is between them. One may extend this concept a bit by considering whether it may make sense to do cross-spectral analysis even in the absence of peaks in the power spectrum. Suppose we have two time series whose power spectra both are indistinguishable from red noise? Under these circumstances what might cross-spectral analysis still be able to reveal? It might be that within this red noise spectrum there are in fact coherent modes at particular frequencies. We can test for this by looking at the coherency spectrum.

7.15.1 Complex Fourier Transform of Cross-Spectrum

The formulation for cross-spectral is based on Fourier Analysis. It is most compact to express this in complex form, where variables have a real and an imaginary component. Any time series on the interval $0 < t < T$ can be expressed as a complex Fourier transform.

$$x(t) = \sum_{k=-N/2}^{N/2} F_x^k e^{i\frac{2\pi k t}{T}} \qquad y(t) = \sum_{k=-N/2}^{N/2} F_y^k e^{i\frac{2\pi k t}{T}} \tag{7.60}$$

The cross spectrum of x and y is performed by taking the product of the Fourier coefficients.

$$C_{xy}^k = F_x^k F_y^{k*} \tag{7.61}$$

where the asterisk indicates a complex conjugate. The complex Fourier coefficients can be written in polar form.

$$F_x^k = F_x^{kRe} + iF_x^{kIm} = R_x^k e^{i\theta_{xk}} \tag{7.62}$$

where

$$R_x^k = ((F_x^{kRe})^2 + (F_x^{kIm})^2)^{1/2} \tag{7.63}$$

and R_x^k is real.

With these definitions, and integrating over time, we find that the cross spectrum of x and y is,

$$C_{xy}^k = R_x^k R_y^k e^{i(\theta_x^k - \theta_y^k)} = R_x^k R_y^k e^{i\Delta\theta^k} = R_x^k R_y^k (\cos\Delta\Theta^k + i\sin\Delta\Theta^k) = Co^k + iQ^k \tag{7.64}$$

We thus see that the cross spectrum has a real part, the cospectrum, Co, and an imaginary part the quadrature spectrum, Q, whose ratio determines the phase difference between the two time series at each frequency indicated by the index k.

$$\Delta\theta^k = \arctan Q^k / Co^k \tag{7.65}$$

The cospectrum and quadrature spectrum can be averaged over many realizations and also over frequency to compute robust phase differences between time series. The coherence uses the averaged spectra and cross spectra to measure the degree to which the phase and amplitude differences remain constant across the realizations and frequencies that are averaged over. The coherence-squared is expressed in terms of the averaged Power spectra of the two time series, Φ_x^k and Φ_y^k and the averaged co-spectra and quadrature spectra squared.

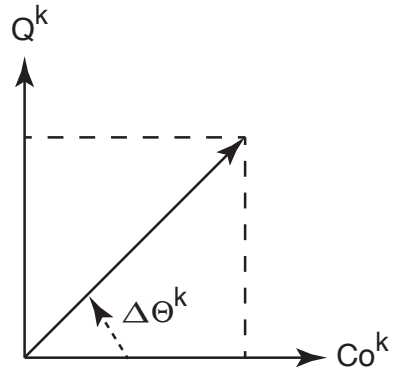

Figure 7.14 Diagram showing the relationship between the co-spectrum Co^k, quadrature spectrum Q^k and the phase angle between two time series $\Delta\Theta^k$ as a function of frequency index, k.

$$\mathrm{Coh}^2(k) = \frac{\mathrm{Co}(k)^2 + \mathrm{Q}(k)^2}{\Phi(k)\Phi(k)} \tag{7.66}$$

The coherence is analogous to the correlation coefficient, except it is a function of frequency. For a single realization of a single frequency the coherency is one. As different realizations are averaged together, the coherency will decline if the phase difference or amplitude ratio of the two time series varies from realization to realization. Coherence measures the consistency of the linear relationship between the two time series as the number of realizations is increased. Tables for the statistical significance of the coherence have been prepared that can test the null hypothesis that the coherence is zero (Amos and Koopmans, 1963) 10.1. The uncertainty in the phase difference can also be related to the coherence (Goodman, 1957). The uncertainty in the phase difference estimation increases as the coherence is reduced (Hartmann, 1974).

7.15.2 Example: Rossby-Gravity Wave Cross-Spectral Analysis

In Figure 7.13d it is shown that the ratio of meridional wind variance in the 3 to 6 day period band in the central equatorial Pacific peaks in the lower troposphere on the equator. This is the structure we expect for the Mixed Rossby-Gravity wave. In this section we perform cross-spectral analysis between the meridional wind at the point on the equator at 210E and 850hPa and the meridional wind at other latitudes, longitudes and pressures. Figure 7.15 shows coherence and phase as functions of both longitude and latitude and pressure and latitude. The spectra are averaged over many realizations and all frequencies corresponding to periods between 3.5 and 5 days. This analysis includes 91 realizations of its 64-day chunk length, and 7 frequencies are averaged together, so the cross-spectral analysis has about 600 degrees of freedom. The coherence has statistically significant and reasonably large values over most of the domain shown. The phase decreases toward the west and upward, indicating waves with westward and upward phase movement. The waves move westward with time and tilt eastward with height in the troposphere. In the stratosphere the waves tilt the opposite direction, consistent with upward propagation there (Holton and Hakim, 2012). They change phase by about 90-degrees in 20-degrees of longitude and so have an effective wavelength of about 8,000km, and move westward at about $10 \mathrm{ms}^{-1}$.

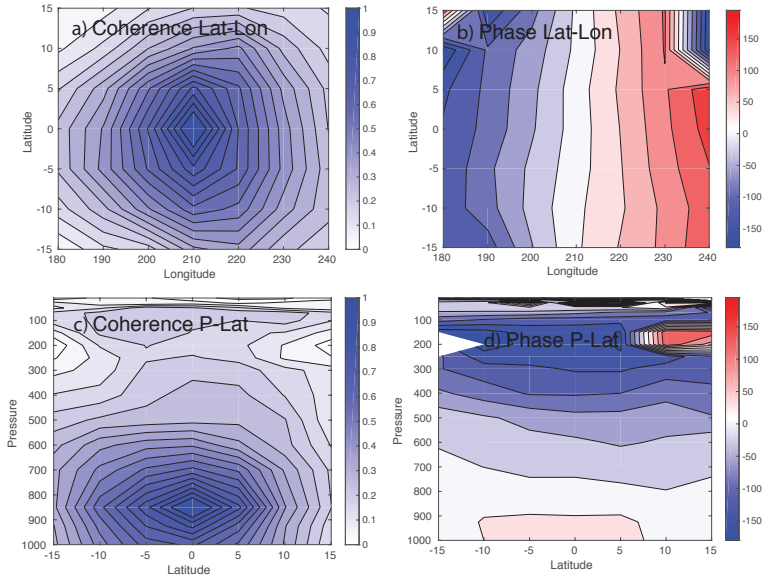

Figure 7.15 Cross-spectral analysis of the meridional wind at the equator, 850hPa and 210E, for periods in the range of 3.5 to 5 days. a) Coherence with the meridional wind at other latitudes and longitudes at 850hPa. b) Phase difference with other latitudes and longitudes at 850hPa. c) Coherence with the meridional wind at other pressures and latitudes. d) Phase difference with the meridional wind at other pressures and latitudes. A chunk length of $M = 64$ was used. Contour interval for coherence is 0.05 and for phase is 30 degrees.

7.16 Space-Time Spectrum Analysis

Mixed space-time spectral analysis is a straightforward extension of harmonic analysis to two dimensions. It is most convenient if the spatial dimension is cyclically continuous, such as in the case of latitude circles, or at least that the spatial dimension has fixed boundaries, like an ocean basin. In such cases we can look for modes of variability in which spatial scales have particular temporal scales. If the behavior is indeed harmonic (wavelike), then we expect mixed space-time spectral analysis to isolate any such modes that are present. For example, if one did mixed space-time spectral analysis of a stringed instrument, one would definitely expect to find a definite relationship between the length scales and the time scales of the oscillations.

Suppose we have a function of longitude, λ , and time, t. We can write:

$$x(\lambda, t) = \sum_{k} \sum_{\pm\omega} X_{k,\pm\omega} \cos(k\lambda \pm \omega t + \Theta_{k,\pm\omega}) \tag{7.67}$$

where $+\omega$ and $-\omega$ correspond to eastward- and westward-moving waves, respectively (Hayashi, 1971, 1979). The zonal wavenumber, k, is the number of zero crossings of the cosine wave around a latitude circle.

If we have such an expansion then we can write the power spectrum as function of both wavenumber and frequency as,

$$P_{k,\pm\omega}(x) = \frac{1}{2} X^2_{k,\pm\omega} \tag{7.68}$$

If we have two time series $x(\lambda, t)$ and $y(\lambda, t)$ we can write the cospectrum between x and y as,

$$Co_{k,\pm\omega}(x, y) = \frac{1}{2} X_{k,\pm\omega} Y_{k,\pm\omega} \cos\big(\Theta_{k,\pm\omega}(y) - \Theta_{k,\pm\omega}(x)\big) \tag{7.69}$$

and the quadrature spectrum is,

$$Q_{k,\pm\omega}(x, y) = \frac{1}{2} X_{k,\pm\omega} Y_{k,\pm\omega} \sin\big(\Theta_{k,\pm\omega}(y) - \Theta_{k,\pm\omega}(x)\big) \tag{7.70}$$

And so the coherence-squared is written,

$$Coh^2_{k,\pm\omega}(x, y) = \frac{Co^2_{k,\pm\omega}(x, y) + Q^2_{k,\pm\omega}(x, y)}{P_{k,\pm\omega}(x) \cdot P_{k,\pm\omega}(y)} \tag{7.71}$$

To obtain the expansion 7.67 we proceed by first performing a Fourier analysis in the longitude coordinate.

$$x(\lambda, t) = \sum_{k} C_k(t) \cos(k\lambda) + S_k(t) \sin(k\lambda) \tag{7.72}$$

We then perform a Fourier analysis in time of these cosine and sine coefficent time series.

$$\begin{aligned} C_k(t) &= \sum_{\omega} A_{k,\omega} \cos(\omega t) + B_{k,\omega} \sin(\omega t) \\ S_k(t) &= \sum_{\omega} a_{k,\omega} \cos(\omega t) + b_{k,\omega} \sin(\omega t) \end{aligned} \tag{7.73}$$

Hayashi (Hayashi, 1971) shows through a straightforward manipulation that A,B,a and b can be related to $X_{k,\pm\omega}$ as follows.

$$4X^2_{k,\pm\omega} = (A \mp b)^2 + (\pm B - a)^2 \tag{7.74}$$

where $X^2_{k,\pm\omega}$ is the space-time power spectrum we desire and the phase is given by,

$$\phi_{k,\pm\omega} = \tan^{-1}\Big(\frac{\mp B - a}{A \pm b}\Big) \tag{7.75}$$

7.16.1 Standing Waves

In space-time spectral analysis standing waves appear as equal contributions from eastward and westward propagating waves. For example, consider the following simple analytic wave.

$$\begin{aligned} x(\lambda, t) &= \cos(k\lambda)\cos(\omega t) \\ &= \frac{1}{2}\big(\cos(k\lambda - \omega t) + \cos(k\lambda + \omega t)\big) \end{aligned} \tag{7.76}$$

We thus see that a stationary wave whose amplitude oscillates in time is equivalent to equal amplitude eastward and westward propagating waves with the same wavenumber and frequency. If we do mixed space-time spectral analysis of such a wave it is hard to tell if it is two independent waves traveling eastward and westward with the same wavenumber and frequency, or a stationary wave whose amplitude oscillates in time. Figure 7.16 shows the space-time spectrum of an analytic zonal wavenumber 5 with a period of 5 days computed with a window of 32 daily observations and based on a record of 264 days. Other details of the procedure are discussed in section 7.16.2.

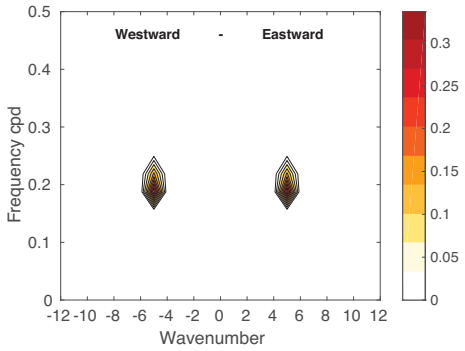

Figure 7.16 Space-time spectral analysis of a standing wave of wavenumber 5 and period 5 days. The time spectral analysis was done with a 32-day chunk and based on 264 days of record. Negative wavenumbers indicate westward traveling waves. Frequency is given in cycles per day.

Such a spectrum presents two possibilities; A standing wave or two independent waves traveling in opposite directions. How does one distinguish these two possibilities? One way to approach this problem is to ask if the eastward and westward waves are related, are they coherent with each other, do they bear a constant phase relationship to each other (standing wave), or are the eastward and westward waves linearly independent? Two somewhat different approaches to this question have been presented. One way to judge this is to formulate a coherence-squared between the eastward and westward waves (Pratt (1976); Hayashi (1977, 1979)). Another method is to look at the coherence in time between the sine and cosine coefficients of a particular wavenumber. Schäfer (1979) uses the coherence in time of the sine and cosine coefficients to ask whether what is seen are "waves" or "noise".

An alternative method of separating standing and traveling oscillations has been proposed by Watt-Meyer and Kushner (2015). This method assumes that the standing wave is the minimum amplitude of the eastward and westward components, and defines the traveling component as the difference between the actual value and the minimum value.

$$X^{Standing}_{k,\pm\omega} = \min\Big(X_{k,+\omega}, X_{k,-\omega}\Big) \tag{7.77}$$

$$X^{Traveling}_{k,\pm\omega} = X_{k,\pm\omega} - X^{Standing}_{k,\pm\omega} \tag{7.78}$$

Here $X_{k,\pm\omega}$ is defined in 7.67. Using the definitions in 7.77 and 7.78 has the advantages that the standing and traveling components can be reconstructed for display, and the possibly important covariance between the standing and traveling components can be explicitly computed.

7.16.2 Example of Space-Time Spectral Analysis

In this section we will discuss an example of space-time spectral analysis of tropical winds and outgoing longwave radiation (OLR). We obtain wind data from the ERA-Interim reanalysis project (Dee et al., 2011) and OLR data from the NOAA daily record (Liebmann and Smith, 1996). Daily instantaneous 0Z ERA Interim data were used beginning in 1979 and extending for 16 years. The daily OLR data used extend from 1979 through 2013.

Prior to spectral analysis the data were processed in the following way. The annual cycle was first removed by fitting the first 4 harmonics of the annual period (365.25 days) to the entire data set. Next The data were averaged in latitude from 15S to 15N to measure the equatorially symmetric part of the variability, and a difference was taken of the average from the equator to 15N minus the average of the equator to 15S to measure the equatorially anti-symmetric part of the variability. At this point we have time series at each longitude. Next we prewhiten each time series by removing the red noise using the autocorrelation for each time series following the procedure outlines in section 7.12. This is more objective than using the smoothing and ratioing procedure employed by Wheeler and Kiladis (1999), and it retains the correct ratio of variance at different frequencies. The data are now ready for analysis. The first step is to do a Fourier transform in longitude to divide the variance into different zonal wavenumbers as in 7.72. The data are spaced at 2.5 degrees, so that there are 72 grid points around a latitude circle. Since this is not a power of two, this Fourier Transform is done by regressing sine and cosine functions onto the data in longitude at each time.

Next we need to perform the time Fourier analysis as in 7.73. For this we wish to use a fast Fourier Transform, and therefore need to divide the time series into chunks whose lengths are powers of 2. If we choose a chunk length of 128 days we have 45 realizations in a time series of 16 years. We therefore divide the time series in to chunks of 128 days. Because we are going to use a Hamming filter, we overlap these chunks by 50%. Once we have selected a chunk, we remove the linear trend and multiply it by the Hamming window function. The FFT program we use returns the complex coefficients of a complex Fourier transform, which we wish to translate into the real coefficients of a sine and cosine expansion.

$$\begin{aligned} x(t) &= \sum_{-\omega}^{\omega} F_\omega e^{i\omega t} \\ x(t) &= \sum_{\omega} A_\omega \cos(\omega t) + B_\omega \sin(\omega t) \end{aligned} \tag{7.79}$$

If the original time series $x(t)$ is real then the complex Fourier coefficients corresponding to positive and negative ω must be complex congugates of each other. We can then easily show that,

$$A_\omega = 2\mathrm{Real}(F_\omega) \qquad B_\omega = -2\mathrm{Imag}(F_\omega) \tag{7.80}$$

With these identities we can rationalize the real formulation of (Hayashi, 1971) with a standard complex FFT. We cannot average the Fourier transforms, but must calculate the power spectrum $X^2_{k,\pm\omega}$ for each chunk and then average those together to get our averaged spectrum.

We plot the averaged space-time spectra with zonal wavenumber as the abscissa and frequency as the ordinate, and denote westward moving waves with negative wavenumbers. This has the advantage of connecting the Mixed Rossby-Gravity wave across zero wavenumber in a way that would be less clear if the axes were reversed. Figure 7.17 shows a few plots of the space-time spectra of OLR and wind that illustrate the power of this technique.

Figure 7.17a shows the space-time power spectrum of equatorially symmetric OLR. Westward propagating Rossby waves are seen, but more prominent are the eastward-propagating Madden-Julian Oscillation (MJO) at low frequencies and low zonal wavenumbers and eastward propagating convectively-coupled Kelvin waves spanning wavenumbers 2 to 8 and periods from 10 to 3 days. The anti-symmetric OLR (Fig. 7.17b)shows some low-frequency variability across a larger range of wavenumbers, the westward propagating Rossby waves and a new feature centered at zonal wavenumber zero and a period of 4 days that represents a broad range of mixed Rossby-gravity waves corresponding to those investigated in section 7.12.1.

Space-time power spectra for symmetric 850hPa winds are shown in Figures 7.17c,d. The MJO is very apparent in the symmetric zonal wind, as are the Kelvin waves. The symmetric zonal wind also shows a

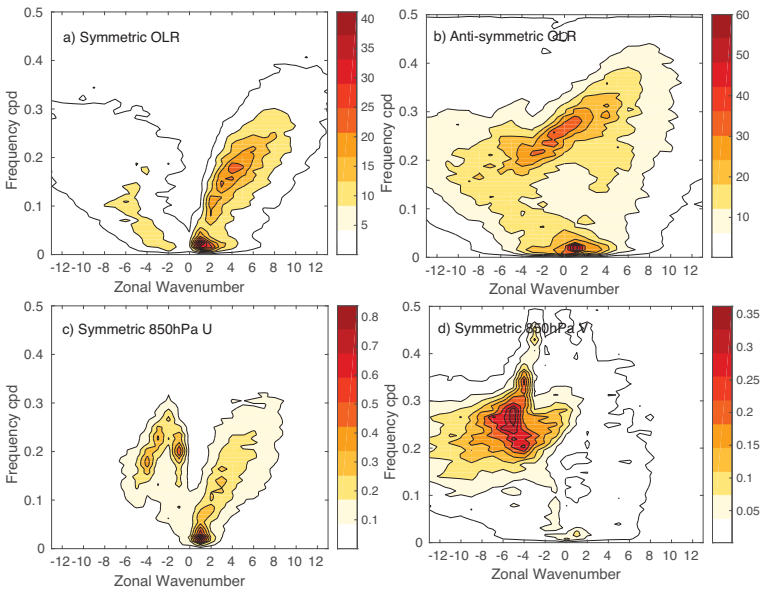

Figure 7.17 Space-time spectral analysis of equatorial waves. a) OLR averaged over 15S-15N, b) OLR averaged over 0-15N minus OLR averaged over 0-15S, c) Zonal wind at 850hPa averaged over 15S-15N, d) Meridional wind averaged over 15S-15N. Negative wavenumbers indicate westward traveling waves. Frequency is given in cycles per day.

nice peak in variance associated with the 5-day wave of zonal wavenumber 1 (Geisler and Dickinson, 1976; Hendon and Wheeler, 2008). The mixed Rossby-gravity waves are seen in the symmetric meridional wind. The westward propagating Rossby waves seen in the OLR show up better in the space-time power spectra of the asymmetric components of velocity, which are not shown here.