6. Time (or Space) Series Analysis

In this section 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 geophysics and in many other fields.

6.1 Autocorrelation

6.1.1 The Autocorrelation Function

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

$$\phi(\tau) = \frac{1}{t_2 - t_1 - \tau} \int_{t_1}^{t_2-\tau} x'(t)x'(t+\tau)dt \tag{6.1}$$

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

$$\phi(L) = \frac{1}{N-2L} \sum_{k=L}^{N-L} x'_k\, x'_{k+L} = \overline{x'_k\, x'_{k+L}};\ L = 0, \pm1, \pm2, \pm3, \ldots \tag{6.2}$$

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 $\phi(\tau)/\phi(0) = r(\tau)$; $-1 \le r(\tau) \le 1$; $r(0) = 1$; if x is not periodic $r(\tau) \to 0$, 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. 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, $\phi(\tau) = \phi(-\tau)$. Under the assumption that the statistics of the data set are stationary in time, it would also be reasonable to extend the summation in (6.2) from $k=L$ to N in the case of negative lags, and from $k=1$ to $N-L$ in the case of positive lags. Such an assumption of stationarity is inherent in much of what follows.

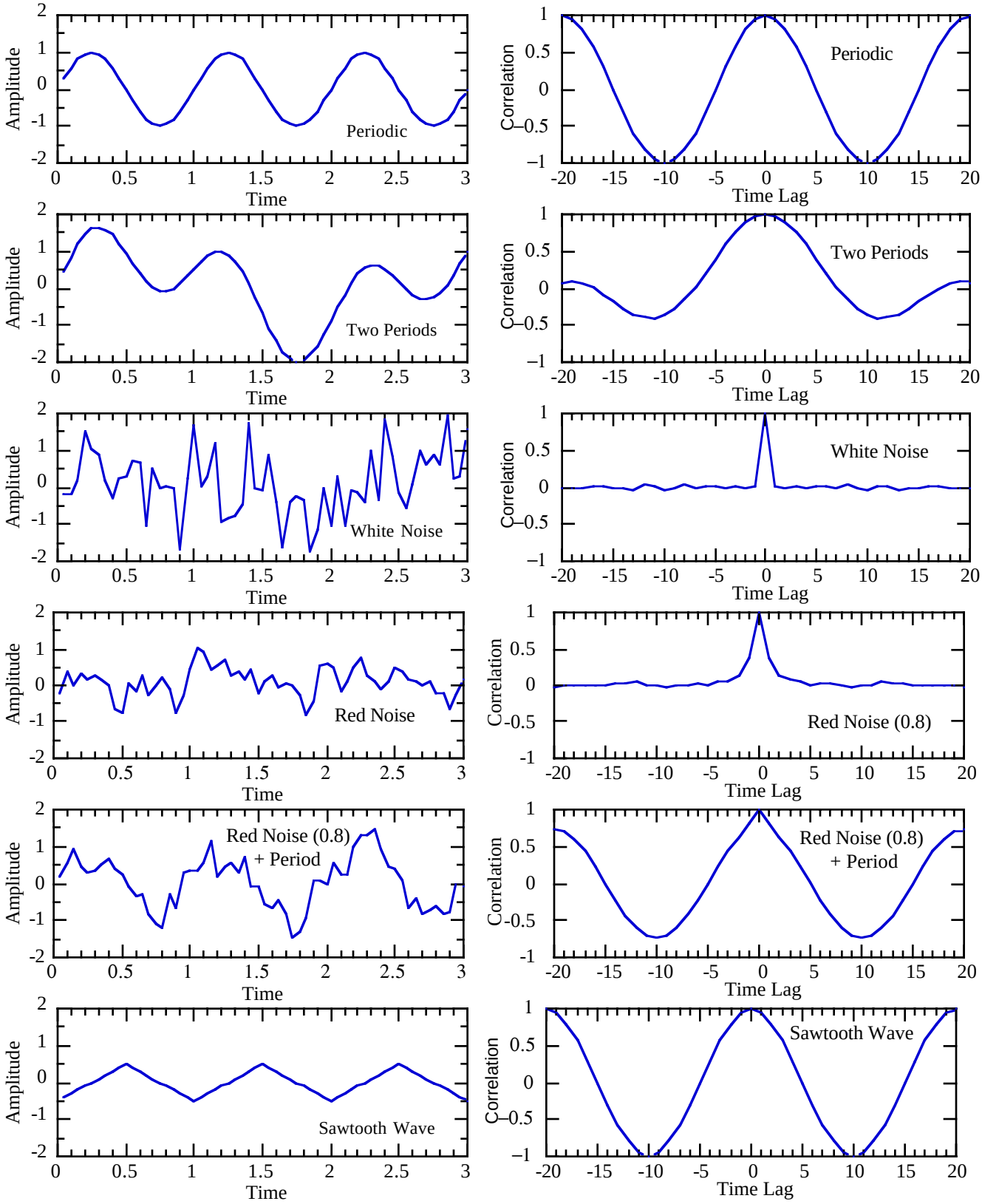


Figure. Comparison of time series and autocorrelation functions for various cases.

6.1.2 Red Noise:

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

$$x(t) = a\,x(t-\Delta t) + (1-a^2)^{1/2}\,\varepsilon(t) \tag{6.3}$$

where x is a standardized variable $\left(\bar{x}=0,\ \overline{x'^2}=1\right)$, a is on the interval between zero and one and measures the degree to which memory of previous states is retained ($0 \le a \le 1$), ε is a random number drawn from a standardized normal distribution, and Δt is the time interval between data points.

Multiply (6.3) by $x(t-\Delta t)$ and average to show that a is the one-lag autocorrelation, or the autocorrelation at one time step, Δt.

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

$$= a \cdot 1 + \left(1-a^2\right)^{1/2} \cdot 0$$

$$\overline{x(t-\Delta t)x(t)} = r(\Delta t) = a$$

; *i.e.*, a is the autocorrelation at Δt

Projecting into the future, we obtain

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

$$= a^2x(t-\Delta t) + a\left(1-a^2\right)^{1/2}\varepsilon + \left(1-a^2\right)^{1/2}\varepsilon$$

$$= a^2x(t-\Delta t) + (a+1)\left(1-a^2\right)^{1/2}\varepsilon$$

Now $x(t+2\Delta t) = ax(t+\Delta t) + \varepsilon$. Consistent with (6.2), multiply by $x(t)$ and average using the definitions above.

$$\overline{x(t)x(t+2\Delta t)} = \overline{ax(t+\Delta t)x(t)} + \overline{\varepsilon x(t)}$$

$$r(2\Delta t) = a\,r(\Delta t) + 0$$

$$r(2\Delta t) = \left(r(\Delta t)\right)^2$$

or by induction

$$r(n\Delta t) = r^n(\Delta t)$$

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

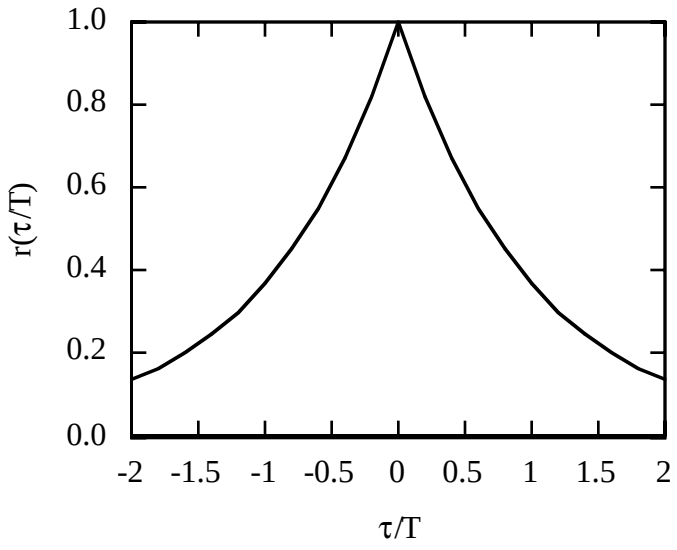
exponential function, $e^{nx} = \left(e^{x}\right)^{n}$, so we may hypothesize that the autocorrelation function for red noise has an exponential shape.

$$r(n\Delta t) = \exp\{-n\Delta t/T\}\,.$$

or if $\tau = n\Delta t$,

$$r(\tau) = \exp(-\tau/T) \quad \text{where T = } -\Delta t/\ln a \tag{6.4}$$

The autocorrelation function for *red noise*.



In summary, if we are given a red noise time series, also called a Markov process,

$$x(t) = ax(t-\Delta t) + \left(1-a^{2}\right)^{1/2}\varepsilon(t) \tag{6.5}$$

then its autocorrelation function is,

$$r(\tau) = \exp(-\tau/T) \tag{6.6}$$

where the autocorrelation e-folding decay time is given by, $\text{T} = -\Delta t/\ln a$

6.1.3 Statistical Prediction and Red Noise

Consider a prediction equation of the form

$$\hat{x}(t+\Delta t) = a'\,x(t) + b'\,x(t-\Delta t) \tag{6.7}$$

where $\bar{x} = 0$. a' and b' are chosen to minimize the rms error on dependent data. Recall from our discussion of multiple regression that for two predictors x_1 and x_2 used to predict y

$$\left|r(x_2, y)\right| \geq \left|r(x_1, y) r(x_1, x_2)\right|$$

In the case where the equality holds, $r(x_2,y)$ is equal to the "minimum useful correlation" discussed in Chapter 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

$$r(2\Delta t) > r(\Delta t)^2$$

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

$$r(2\Delta t) = r(\Delta t)^2$$

so that the value at two lags previous to now always contributes exactly the minimum useful, and nearly automatic, correlation, and there is no point in using a second predictor if the variable we are trying to predict is red noise. All we can use productively is the present value and the autocorrelation function,

$$x(t + \Delta t) = x(t) \quad \text{with an } R^2 = a^2 = r(\Delta t)^2$$

This is just what is called a persistence forecast, we assume tomorrow will be like today.

6.1.4 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, even the present value is of no help in projecting into the future. The probability density function we use is generally normally distributed about zero mean, and this is generated by the 'randn' function in Matlab.

6.1.5 Degrees of Freedom/Independent Samples

Leith [*J. Appl. Meteor.*, 1973, p. 1066] has argued that for a time series of red noise, the number of independent samples N^* is given by

$$N^* = \frac{N\Delta t}{2T} = \frac{\text{total length of record}}{\text{two times } e\text{ - folding time of autocorrelation}} \tag{6.9}$$

where N is the number of data points in the time series, Δt is the time interval between data points and T is the time interval over which the autocorrelation drops to $1/e$. In other words, the number of degrees of freedom we have is only half of the number of e-folding times of

data we have. The more autocorrelated our data is in time, the fewer degrees of freedom we get from each observation.

For Red Noise:

$$r(\tau) = e^{-\tau/T} \quad \ln(r(\tau)) = -\tau/T$$

thus

$$T = -\tau/\ln(r(\tau))$$

e.g., for $\tau = \Delta t$ $T = -\Delta t/\ln[r(\Delta t)]$, so that

$$\frac{N^*}{N} = -\frac{1}{2}\ln[r(\Delta t)] \; ; \frac{N^*}{N} \le 1 \tag{6.10}$$

Sample Values:

$r(\Delta t)$	< 0.16	0.3	0.5	0.7	0.9
N^*/N	1	0.6	0.35	0.18	0.053

Leith's formula (6.9) is consistent with Taylor(1921) for the case of a red noise process. Taylor said that

$$\frac{N^*}{N} = \frac{1}{2L} \tag{6.11}$$

Where L is given by,

$$L = \int_0^\infty r(\tau')\, d\tau' \tag{6.12}$$

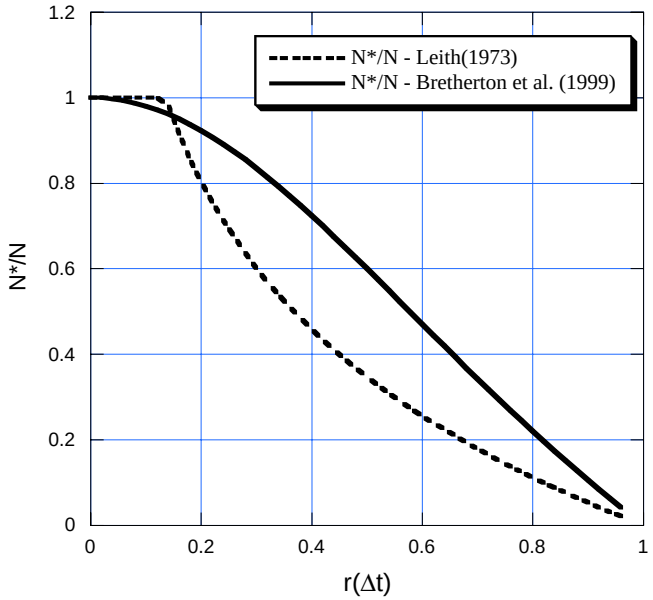
If we substitute the fomula for the autocorrelation function of red noise, (6.4) into 6.12), then we get that L=T, and Taylor's formulat is the same as Leith's. You may see a dimensional inconsistency in (6.11), but this disappears if you consider that Taylor is using time in nondimensional units of the time step, t'=t/Δt, τ'=τ/Δt, so that L=T/Δt.

The factor of two comes into the bottom of the above expression 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 by 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 points more can be explained. This may seem a bit conservative.

Indeed, Bretherton et al, (1999) suggest that, particularly for variance and covariance analysis a more accurate formula to use is:

$$\frac{N^*}{N} = \frac{\left(1 - r(\Delta t)^2\right)}{\left(1 + r(\Delta t)^2\right)} \tag{6.13}$$

If we compare the functional dependence of N*/N from Bretherton et al.(1999), formula (6.13) with that of Leith/Taylor from formula (6.10) we can make the plot below.



So you can see that the Bretherton et al. formula, which is appropriate for use in covariance problems, is more generous than Leith's conservative formula, allowing about twice as many degrees of freedom when the autocorrelation at one lag is large.

References:

Bretherton, C. S., M. Widmann, V. P. Dymnikov, J. M. Wallace and I. Blade, 1999: The effective number of spatial degrees of freedom of a time-varying field. *J. Climate*, **12**, 1990-2009.

Leith, C. E., 1973: The standard error of time-averaged estimates of climatic means. *J. Appl. Meteorol.*, **12**, 1066-1069.

Taylor, G. I., 1921: Diffusion by continuous movement. *Proc. London Math. Soc.*, **21**, 196-212.

6.1.6 Verification of Forecast Models

Consider a forecast model that produces a large number of forecasts x_f of x. The mean square (ms) error is given by

$$\overline{\left(x - x_f\right)^2}$$

The skill of the model is related to the ratio of the ms error to the variance of x about its climatological mean. Suppose that the model is able to reproduce climatological statistics in the sense that

$$\bar{x}_f = \bar{x}, \overline{x'^2_f} = \overline{x'^2}.$$

If the model has no skill then

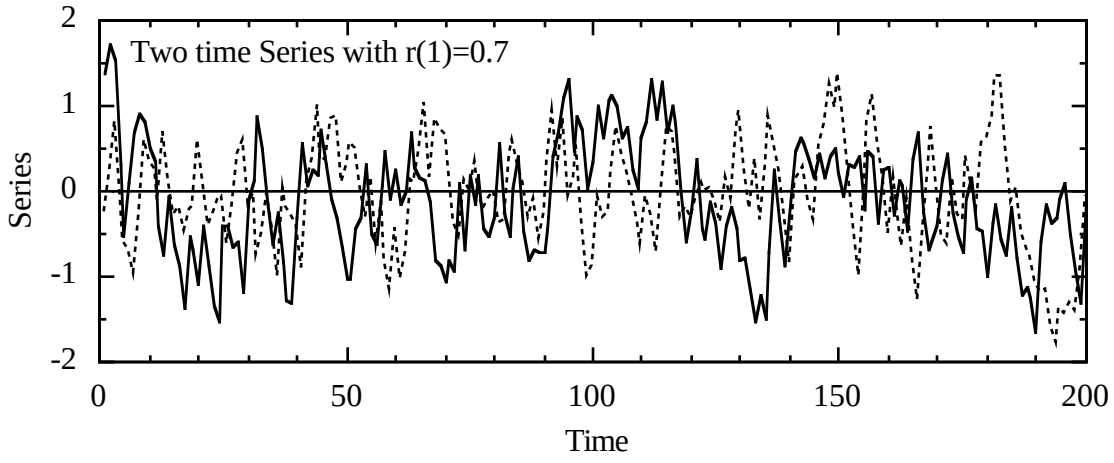
$$\overline{x' x'_f} = 0$$

so that

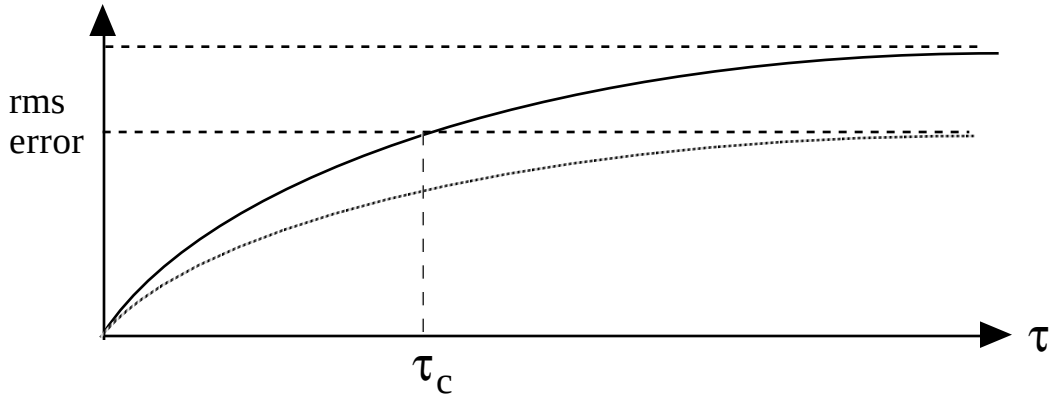
$$\overline{\left(x - x_f\right)^2} = \overline{\left(x' - x'_f\right)^2} = \overline{x'^2} - 2\overline{x' x'_f} + \overline{x'^2_f} = 2\overline{x'^2}$$

This result may seem somewhat paradoxical at first. Why is it not simply $\overline{x'^2}$?

The average root mean squared difference between two randomly chosen values with the same mean is larger by √2' than that of each of these values about their common mean.



The following figure shows a plot of the rms error versus prediction time interval τ for a hypothetical forecast model whose skill deteriorates to zero as $\tau \to \infty$.



For $\tau > \tau_c$ the model appears to have no skill relative to climatology, yet it is clear that it must still have some skill in an absolute sense since the error has not yet leveled off. The model can be made to produce a forecast superior to climatology if we use a regression equation of the form.

$$\hat{x}_f = a x_f + (1-a)\bar{x}$$

Least-squares regression to minimize

$$\left(x - \hat{x}_f\right)^2$$

yields

$$a = \frac{\overline{x' x'_f}}{\overline{x'^2}} \equiv \text{The multiple correlation factor, } R, \text{ for the original regression}$$

So we should choose $a = R(\tau)$.

As the skill of the prediction scheme approaches zero for large τ the $\hat{x}_f$ forecast is weighted more and more heavily toward climatology and produces an error growth like the dotted curve in the figure above.

Problem:

Prove that at the point where the rms error of a simple forecast x_f, passes the error of climatology (the average), where $\tau = \tau_c$, $a = 0.5$, and at that point the rms error of $\hat{x}_f$ equals 0.87 times the rms error of x_f.

6.2 Harmonic Analysis

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 we have a set of N values of $y(t_i) = y_i$. Then we can use a least-squares procedure to find the coefficients of the following expansion

$$y(t) = A_o + \sum_{k=1}^{\frac{N}{2}} \left(A_k \cos 2\pi k \frac{t}{T} + B_k \sin 2\pi k \frac{t}{T} \right) \tag{6.14}$$

where T = the length of the period of record. $y(t)$ is a continuous function of t which passes through all of the data points. Note that $B_k = 0$ when $k=N/2$, since you cannot determine the phase of the wave with a wavelength of two time steps. If the data points are *not evenly spaced* in time then we must be careful. The results can be very sensitive to small changes in y_i. One should test for the effects of this sensitivity by imposing small variations in y_i and be particularly careful where there are large gaps. Where the data are unevenly spaced it may be better to eliminate the higher harmonics. In this case one no longer achieves an exact fit, but the behavior may be much better between the data points.

Useful Math Identities:

It may be of use to you to have this reference list of trigonometric identities:

$$\cos(\alpha - \beta) = \cos\alpha \cos\beta + \sin\alpha \sin\beta \quad ; \quad \tan\gamma = \frac{\sin\gamma}{\cos\gamma} \tag{6.15}$$

You can use the above two relations to show that:

$$C\cos\theta + S\sin\theta = A\cos(\theta - \theta_o); \quad where\ A = \sqrt{C^2 + S^2} \ ; \ \theta_o = Arc\tan\left(\frac{S}{C}\right) \tag{6.16}$$

where you need to note the signs of S and C to get the phase in the correct quadrant.
The complex forms of the trig functions also come up importantly here.

$$e^{i\theta} = \cos\theta + i\sin\theta \quad where \ \ i = \sqrt{-1} \tag{6.17}$$

Also, from this you can get;

$$\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} \ ; \quad \cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} \tag{6.18}$$

If you need more of these, check any book of standard mathematical tables.

6.2.1 Evenly Spaced Data → Discrete Fourier Transform

On the interval $0 \le t \le T$ chosen such that t_1=0 and $t_{N+1} = T$ where N is an even number. The analytic functions are of the form

$$\cos(2\pi k\ i\Delta t/T), \sin(2\pi k\ i\Delta t/T) \tag{6.19}$$

Δt is the (constant) spacing between the grid points. In the case of evenly spaced data we have:

a) a_o = the average of y on interval $0 \le t \le T$.

b) The functions (predictors) are orthogonal on the interval $0 \le t \le T$ so that the covariance matrix is diagonal and the coefficients can be determined one at a time (see Section 4.2). Hence

$$a_k = \overline{x'_k y'} \Big/ \overline{x'^2_k} \tag{6.20}$$

c) The functions each have a variance

$$\overline{x'^2_k} = \frac{1}{2}$$

except for $A_{N/2}$ and $B_{N/2}$ whose variances are 1 and 0 respectively.

Hence we derive the rather simple algebraic formulas for the coefficients:

$$A_k = \frac{2}{N}\sum_{i=1}^{N} y_i \cos 2\pi k i\Delta t/T$$

$$B_k = \frac{2}{N}\sum_{i-1}^{N} y_i \sin 2\pi k i\Delta t/T$$

$$k = 1, N_2 - 1$$

$$A_{N/2} = \frac{1}{N}\sum_{i=1}^{N} y_i \cos \pi N i\Delta t/T$$

$$a_o = \frac{1}{N}\sum y_i$$

$$b_o = 0 \tag{6.21}$$

or

$$y(t) = \bar{y} + \sum_{k=1}^{\frac{N}{2}-1} \left\{ A_k \cos\left(2\pi k \frac{t}{T}\right) + B_k \sin\left(2\pi k \frac{t}{T}\right) \right\} + A_{N/2} \cos\left(\pi N \frac{t}{T}\right) \tag{6.22}$$

or, alternatively

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

$$C_k^2 = A_k^2 + B_k^2 \text{ and } t_k = \frac{T}{2\pi k}\tan^{-1}\left(\frac{B_k}{A_k}\right) \qquad (6.23)$$

Of course, normally these formulas would be obtained analytically using the *a priori* information that equally spaced data on a finite interval can be used to exactly calculate the Fourier representation on that interval (assuming cyclic continuation *ad infinitum* in both directions).
The fraction of the variance explained by a particular function is given by

$$r^2(y, x_k) = \frac{\overline{x'_k y'}^2}{\overline{x'^2_k}\;\overline{y'^2}}$$

$$= \frac{A_k{}^2 + B_k{}^2}{2\overline{y'^2}}, \text{for } k = 1, 2, \ldots \frac{N}{2} - 1$$

$$\frac{A_{N/2}^2}{\overline{y'^2}} \text{ for } k = \frac{N}{2} \qquad (6.24)$$

The variance explained by a particular k is

$$\frac{C_k^2}{2} \text{ for } k = 1, 2, \ldots \frac{N}{2} - 1; \;\; A_{N/2}^2 \text{ for } k = \frac{N}{2} \qquad (6.25)$$

6.2.2 The Power Spectrum

The plot of C_k^2 vs. k is called the power spectrum of $y(t)$ - the frequency spectrum if t represents time and the wavenumber spectrum if t represents distance. Strictly speaking C_k^2 represents a line spectrum since it is defined only for integral values of k, which correspond to particular frequencies or wavenumbers. If we are sampling a finite data record from a larger time series, then this line spectrum has serious drawbacks.

1. Integral values of k do not have any special significance, but are simply determined by the length of the data record T, which is usually chosen on the basis of what is available.

$$k_j = j\frac{\Delta t}{T}, j = 0, 1, 2, \ldots N/2$$

2. The individual spectral lines each contain only about 2 degrees of freedom, since N data points were used to determine a mean, $N/2$ amplitudes and $N/2 - 1$ phases (a mean and $N/2$ variances). Hence, assuming a reasonable amount of noise is present, a line spectrum may (should) have very poor reproducibility from one finite sampling interval to another; even if the series is stationary

(i.e., its true properties do not change in time). To obtain reproducible, statistically significant results we need to obtain spectral estimates with many degrees of freedom.

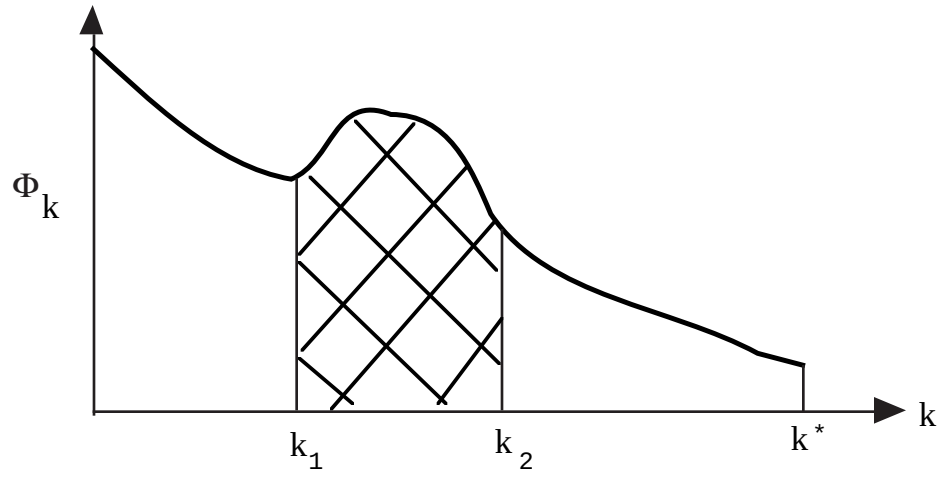
3. With the notable exceptions of the annual and diurnal cycles and their higher harmonics, most interesting "signals" in geophysical data are not truly periodic but only quasi-periodic in character, and are thus better represented by spectral bands of finite width, rather than by spectral lines.

Continuous Power Spectrum: $\Phi(k)$

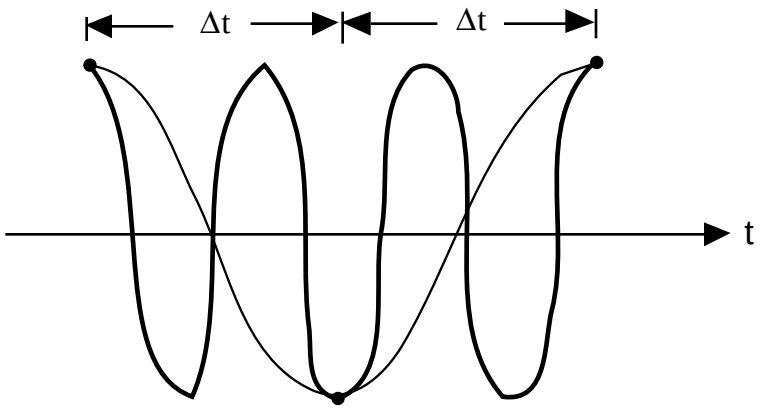
All of the above considerations suggest the utility of a continuous power spectrum which represents the variance of $y_i(t)$ per unit frequency (or wavenumber) interval such that

$$\overline{y'^2} = \int_0^{k^*} \Phi(k)\, dk \tag{6.26}$$

So that the variance contributed is equal to the area under the curve $\Phi(k)$, as shown below.



k^* corresponds to one cycle per $2\Delta t$, the highest frequency in $y(t)$ that can be resolved with the given spacing of the data points. This k^* is called the Nyquist frequency. If higher frequencies are present in the data set they will be aliased into lower frequencies.
$0 \le k \le k^*$



true period $t = 2/3\ \Delta t$ aliased period $t = 2.0\Delta t$

true wavenumber $k = 3.0k_*$ aliased wavenumber $k = 1.0k_*$

This is a problem when there is a comparatively large amount of variance beyond k_*.

Degrees of Freedom—Resolution Tradeoff:

For a fixed length of record, we must balance the number of degrees of freedom for each spectral estimate against the resolution of our spectrum. We increase the degrees of freedom by increasing the bandwidth of our estimates. Smoothing the spectrum means that we have fewer independent estimates but greater statistical confidence in the estimate we retain.

High resolution				Lower resolution
High Information	⇒	Smooth/Average	⇒	Lower Information
Low Quality				High quality in a statistical sense

"Always" insist on adequate quality or you could make a fool of yourself.

The number of degrees of freedom per spectral estimate is given by N/M^* where M^* is the number of independent spectral estimates and N is the actual number of data points $y_i(t)$ regardless of what the autocorrelation is. As long as we use a red-noise fit to the spectrum as our null hypothesis, we don't need to reduce the number of degrees of freedom to account for autocorrelation.

6.2.3 Methods of Computing Power Spectra

Direct Method:

The direct method consists of simply performing a Fourier transform or regression harmonic analysis of $y_i(t)$ to obtain C_k^2. This has become economical because of the Fast Fourier Transform (FFT). Because the transform assumes cyclic continuity, it is desirable to "taper" the ends of the time series $y_i(t)$, as will be discussed in section 6.2.5. When we do a Fourier analysis we get estimates of the power spectrum at $N/2$ frequencies, but each spectral estimate has only two degrees of freedom. A spectrum with so few degrees of freedom is unlikely to be reproducible, so we want to find ways to increase the reliability of each spectral estimate, which is equivalent to a search for ways to increase the number of degrees of freedom of each estimate.

How to obtain more degrees of freedom:

a.) Average adjacent spectral estimates together. Suppose we have a 900 day record. If we do a Fourier analysis then the bandwidth will be 1/900 day^{-1}, and each of the 450 spectral estimates will have 2 degrees of freedom. If we averaged each 10 adjacent estimates together, then the bandwidth will be 1/90 day^{-1} and each estimate will have 20 d.o.f.

b.) Average realizations of the spectra together. Suppose we have 10 time series of 900 days. If we compute spectra for each of these and then average the individual spectral estimates for each frequency over the sample of 10 spectra, then we can derive a spectrum with a bandwidth of 1/900 days^{-1} where each spectral estimate has 20 degrees of freedom.

Lag Correlation Method:

According to a theorem by Norbert Wiener, that we will illustrate below, the autocovariance (or autocorrelation, if we normalize) and the power spectrum are Fourier transforms of each other. So we can obtain the power spectrum by performing harmonic analysis on the lag correlation function on the interval $-T_L \le \tau \le T_L$. The resulting spectrum can be smoothed, or the number of lags can be chosen to achieve the desired frequency resolution. The Fourier transform pair of the continuous spectrum and the continuous lag correlation are shown below.

$$\Phi(k) = \int_{-T_L}^{T_L} r(\tau) e^{-ik\tau} d\tau$$
$$r(\tau) = \frac{1}{2\pi} \int_{-k_*}^{k_*} \Phi(k) e^{ik\tau} dk \qquad (6.27)$$

The maximum number of lags L determines the bandwidth of the spectrum and the number of degrees of freedom associated with each one. The bandwidth is 1 cycle/$2T_L$, and frequencies 0, 1 cycle/$2T_L$, $2/2T_L$, $3/2T_L$, ..., $1/2\Delta t$. There are

$$\frac{\frac{1}{2\Delta t}}{\frac{1}{2T_L}} = \frac{T_L}{\Delta t} \tag{6.28}$$

of these estimates. Each with

$$\left(\frac{T_L / \Delta t}{N} \right)^{-1} = \frac{N}{L} \text{ degrees of freedom.} \tag{6.29}$$

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 if you undertake higher order spectral analysis.

