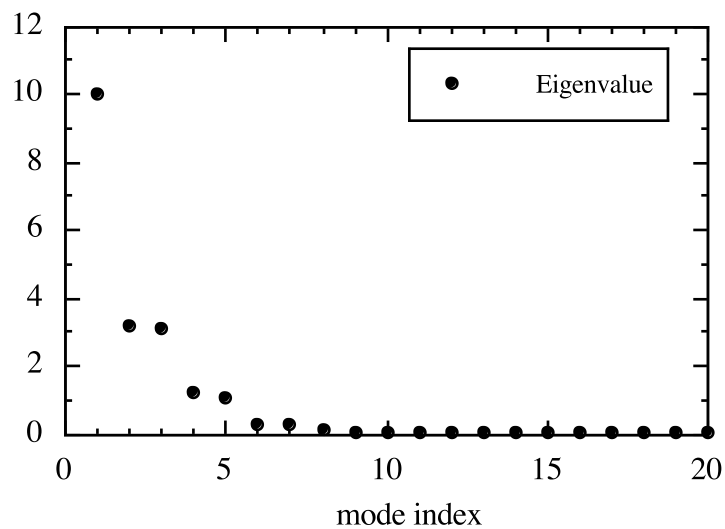
ATMS 552: Objective Analysis Final Preparation

1. A variable y that you wish to predict is correlated with x at 0.7, and with z at 0.6. The two predictors x and z are correlated with each other at the 0.7 level. These correlations are based on a dependent sample of 30 observations. When you apply the regression prediction to independent data in the future, should you use x, z, or both x and z to predict y? Explain your reasoning process.

2. Given the following eigenvalue spectrum, based on a sample of 100 independent realizations, which of the modes of an EOF analysis should you consider as serious contenders for being physical modes of interest, and worthy of further study? Explain your reasoning. Supposing that these modes pass the eigenvalue significance test, what else could you do to test for their physical significance?



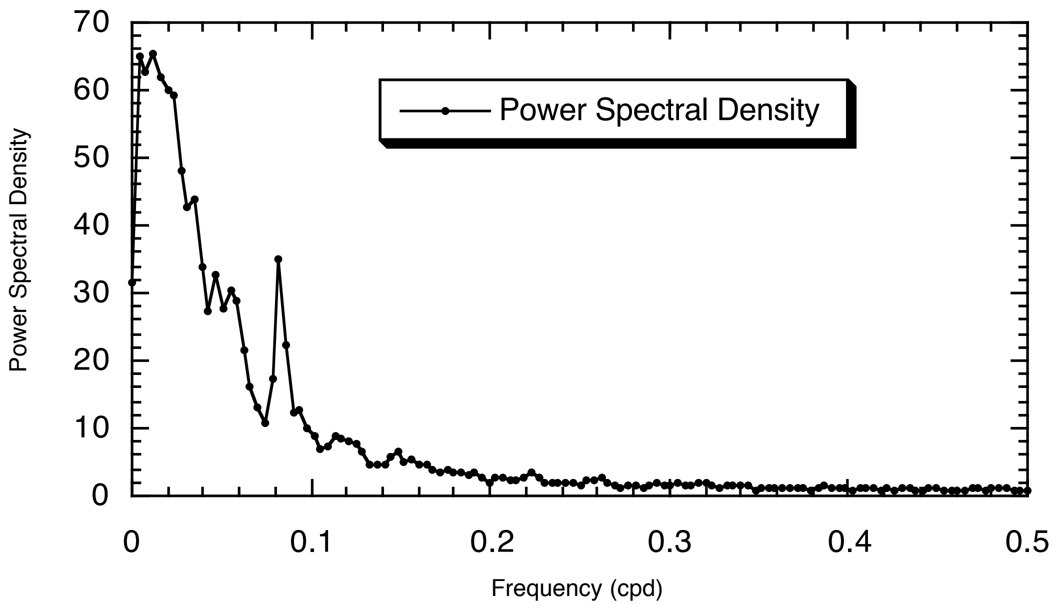
3. You are interested in the Biennial Oscillation(BO) phenomenon of the tropical Pacific Ocean. You want to do a spectral analysis to see whether a 2-year peak appears the power spectrum of sea life in the western tropical Pacific. You might be able to get a long record from a core through a coral head. What sampling interval do you need, and how many years of data do you need if your want to determine the variance as a function of frequency to an uncertainty of 50% and detect a peak at the 95% a priori significance level. You expect that a huge annual cycle is also present. Explain all steps in your reasoning.

4. You want to perform an SVD analysis of an input data set consisting of monthly mean sea surface temperatures(SST) and monthly mean surface pressure for a period of 30 years. The data have been area-averaged over 2˚x2˚ grid squares over the Northern Hemisphere. You have data for about 8,100 pressure grid squares and a slightly

smaller number of SST grids at about 360 times. The pressure data and SST data show a lot of autocorrelation in space and some in time. You want to find the patterns in the atmospheric pressure and the SST that are highly correlated.

Describe the steps you would follow in performing this analysis. In particular your discussion should include, but not be limited to, plans to deal with the annual cycle, the fact that the data are noisy, the fact that you have a huge amount of data and the computations might be tough to do, etc.

5. You are reviewing a paper in which the following power spectrum appears. The spectrum has a bandwith of 0.005 cycles per day and was computed from 1000 daily observations. Decide whether you think the apparent spectral peak is statistically significant. Explain all your reasoning and calculations. You may assume that you have an a priori reason for expecting a feature near a 12-day period. The experimental autocorrelation at one day lag is 0.8. You may sketch on the graph below. You may be approximate in formulating your null hypothesis to minimize calculation (i.e. draw it using an optical function fit, and pencil-hand-eye coordination rather than your calculator.).



6. Compute the response functions, R(ω) for the following 1-2-4-2-1 and 1-3-4-3-1 symmetric nonrecursive filters, using the Fourier Transform method and the time-shifting theorem. Write down the final formulas for the response functions and draw a box around them.

$$\tilde{y}_n = \frac{1}{10}x_{n+2} + \frac{2}{10}x_{n+1} + \frac{4}{10}x_n + \frac{2}{10}x_{n-1} + \frac{1}{10}x_{n-2}$$

$$\tilde{y}_n = \frac{1}{12}x_{n+2} + \frac{3}{12}x_{n+1} + \frac{4}{12}x_n + \frac{3}{12}x_{n-1} + \frac{1}{12}x_{n-2}$$

Sketch the response functions on the Nyquist interval. If your objective is to remove the noise in the highest fifth of the Nyquist interval of frequencies, which of these smoothing schemes is best? You will find the hints useful.

7. Compute the frequency response function for the Haar or Daub2 wavelet and scaling functions.

8. Compute the response function, R(ω) for the following 1-3-1 filter,

$$\tilde{y}_n = \frac{1}{5}x_{n+1} + \frac{3}{5}x_n + \frac{1}{5}x_{n-1}$$

Sketch the response function on the Nyquist interval and compare it with the response function for the 1-2-1 filter. Why do you suppose that the 1-2-1 filter is more often used than the 1-3-1 filter? You will find the following identity useful.

$$\cos x = \frac{e^{ix} + e^{-ix}}{2};$$

9. In performing a spectral analysis searching for a peak at 5 days for which you have *a priori* expectation, you find that the spectrum passes a 99% significance test when the spectrum is computed with a window length of 512 and overlap of 256, but that when you plot the spectrum for a window of 64 with an overlap of 32 the peak at 5 days is not significant at 99%. You have a huge amount of data, and this happens for all subsets and shifts, so you are pretty sure it is in the data and not your analysis technique. . How do you interpret this result?

10. You are interested in the El Niño-Southern Oscillation(ENSO) phenomenon of the tropical Pacific Ocean, that has a meaningful impact, you think, on NW fisheries. You know that warm events recur every 2 to 7 years. You want to do a spectral analysis to see what the power spectrum of an ENSO index is. You might be able to get a long record from coral head data from tropical Pacific reefs. What sampling interval do you need, and how many years of data do you need if your want to determine the variance as a function of frequency to an uncertainty of 50% at the 95% a priori significance level. Explain all steps in your reasoning.

11. You want to perform an EOF analysis of an input data set consisting of climatological monthly mean temperatures (Jan., Feb., ... Dec. where the data have been averaged over all the available years.) that have been mass-averaged through the depth of the Pacific Ocean for grid squares that are about 1˚x1˚ latitude-longitude

squares over the whole ocean. You have data for about 18,000 of these grid square for each of the 12 months.

a) Describe how you would set up an EOF analysis for maximum computational efficiency.

b) What kind of weighting scheme, if any, would you use?

c) How many EOF modes would you get and how many of them would you expect to be physically significant?

14. You have a rapid-response sonic anemometer and you want to look at turbulent microstructures with time scales of about 3 seconds. How often do you need to sample? If you want to know the shape of the spectrum between 1Hertz and 0.01Hertz to within an uncertainty of 50%, how long must you take data? You may assume that the time series is stationary and homogeneous.

16. You have a fairly long sequence of radar observations from an atoll in the Pacific where mesocale convective complexes tend to occur. The radar produces usable returns in an annular volume that extends from about 50km to 200km in radius from the atoll and from about 2km to 12km in altitude. Your digital archive contains returns for this annular volume every 30 minutes for an extended period of time (a couple years, say). From an EOF analysis of the vertical structure you find two EOFs that explain a lot of the variance in radar return: one correlates well with precipitating convective towers and the other correlates well with upper-level stratiform clouds and their precipitation. Devise an analysis scheme for looking at relationships between the occurrence, horizontal location, and scale of the convective towers and the stratiform anvil clouds. *A priori*, you expect that the convective towers feed the stratiform anvils, but have a shorter lifetime and smaller scale. The spatial scale of the basic data is about 1km. Describe the basic techniques to be used. How would the information would be organized for viewing or statistical analysis? What hypotheses could be tested?

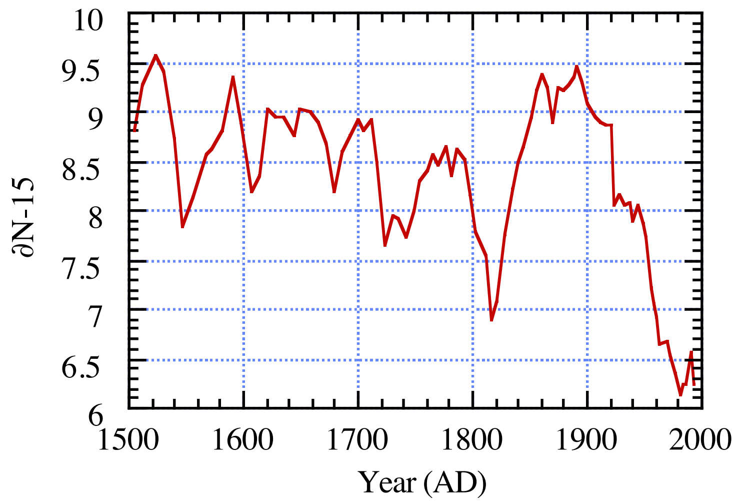
19. What is the impulse response of a recursive filter, and why is it important?

20. Why would you run a recursive filter twice over a time series?

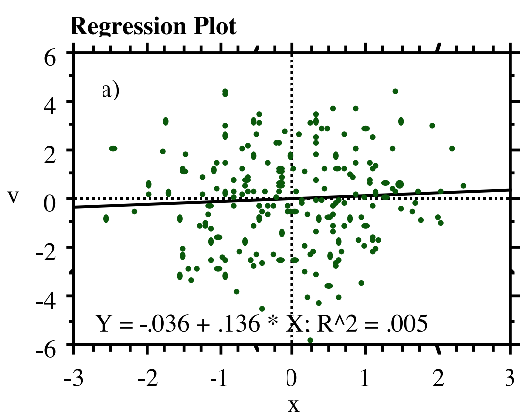
21. Why do you want the matrix operator for a wavelet filter to be orthonormal?

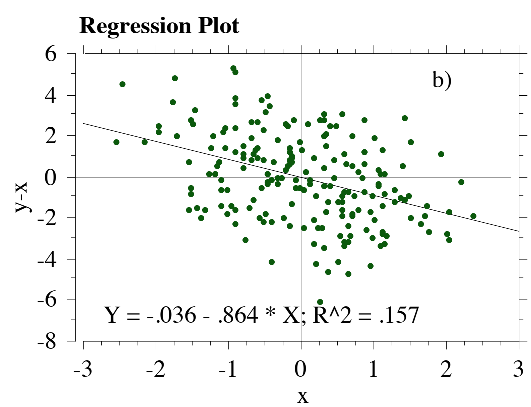
23. Below is a plot of the Nitrogen-15 isotopic abundance in a core taken from the sediment below Karluk Lake, Kodiak, Alaska (B.P.Finney, NPAFC Bull., 1, 388-). The ^{15}N abundance is plotted versus year from about 1500 to about 1992. The ^{15}N abundance is thought to be an indication of the number of sockeye salmon that returned to the streams above Karluk Lake, because ^{15}N is more abundant in Salmon than it is in the other biomass that flows into the sediment.. Assuming that all of this is correct, take a look at the time series from the perspective of a data analyst. What

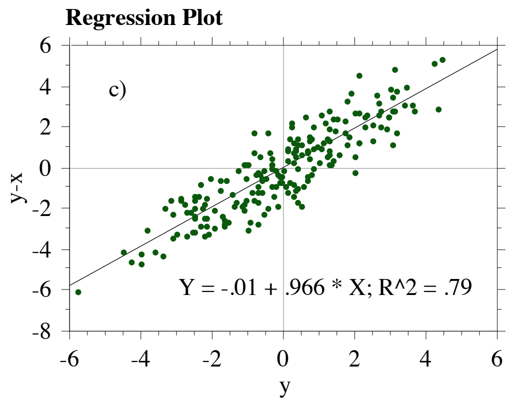
analysis could you conduct with this time series? What techniques might usefully be applied to this time series?



25. Regression: You measure two variables, x and y , at 200 times. If you regress y on x you get the result plotted as a) below. If you plot y-x (y minus x) versus x you get the plot b) below, and if you plot y-x versus y you get the plot c) below. The regression equations for the ordinate versus the abscissa are given in each panel. R^2 is the explained variance of the linear regression. Explain these results as thoroughly as you can. What is the important relationship between x and y? If you can, use mathematics to explain your analysis.






Regression Plot
c)
Y = -.01 + .966 * X; R^2 = .79
y-x
y
6
4
2
0
-2
-4
-6
-8
-6
-4
-2
0
2
4
6

Helpful Hints and Formulae: (But not all are needed.)

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) \quad ; \quad \Delta\lambda = \lambda\sqrt{\frac{2}{N^*}}$$

$$\cos x = \frac{e^{ix} + e^{-ix}}{2}; \quad \sin x = \frac{e^{ix} - e^{-ix}}{2i}; \quad \cos 2x = \cos^2 x - \sin^2 x \ ;$$

$$\sin 3x = 3\sin x - 4\sin^3 x$$

if FT of f(t) = F(ω), then FT of f(t + Δt) = F(ω) $e^{i\omega\Delta t}$

$$\Delta f = \frac{1}{T} \quad ; \quad dof \ge \frac{T}{2} \quad ; \quad f_{Nyquist} = \frac{1}{2\Delta t} \quad or \quad \frac{\pi}{\Delta t} \quad ; \quad R^2 = \frac{r_{1y}^2 + r_{2y}^2 - 2\, r_{1y}\, r_{2y}\, r_{12}}{1 - r_{12}^2}$$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{N-1}}} = \frac{\bar{x} - \mu}{\frac{\hat{s}}{\sqrt{N}}} \ ; \ \hat{s} = \sqrt{\frac{N}{N-1}}\, s$$

$$\cos^2 x = \frac{1}{2}\left(1 + \cos 2x\right), \quad \sin^2 x = \frac{1}{2}\left(1 - \cos 2x\right)$$

$$\mathbf{X} = \mathbf{U\Sigma V}^T \quad \frac{N^*}{N} = \frac{1-\alpha^2}{1+\alpha^2} \qquad R(\omega) = \frac{Y(\omega)}{X(\omega)} = C_0 + 2\sum_{k=1}^{N} C_k \cos\left(\omega k \Delta t\right)$$

$$R(\omega) = \frac{Y(\omega)}{X(\omega)} = \frac{\left\{\sum_{k=0}^{K} a_k\, z^{-k}\right\}}{\left\{1 - \sum_{j=1}^{J} b_j\, z^{-j}\right\}} \quad z = e^{i\omega\Delta t}$$

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)\ e^{i\omega t}\, d\omega \qquad f(t \pm a) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{\pm i\omega a}\, F(\omega)\ e^{i\omega t}\, d\omega$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad \ldots \ P(X = k) = \binom{n}{k} p^k\ (1-p)^{n-k} \ , \ k = 1,2,3,\ldots n$$

$$\text{if } a^n = b \quad then \quad n = \frac{\ln b}{\ln a} \ \ldots \ \lim_{n\to\infty} P\left(a < \frac{X - np}{\sqrt{np(1-p)}} < b\right) = \frac{1}{\sqrt{2\pi np(1-p)}}\int_a^b e^{-x^2/2} dx$$