Homework #6
Objective Analysis in the Atmospheric and Related Sciences
ATMO, HWRS, GEOS, GEOG 529: Fall 2013

Data

Hourly and by minute direct normal solar radiation data (W m^{-2}) from University of Arizona Solar Resource and Meteorological Assessment Project, from 1 October to 20 November, 2013.

Climate indices used in HW #4.

Part I

The following observations of relative humidity at Davis-Monthan AFB (%) were taken every three hours, starting from 6pm, Nov. 19, 2013 (last day is from MOS forecast):

34, 45, 45, 52, 54, 36, 27, 26, 31, 39, 45, 49, 52, 40, 31, 36, 48, 57, 66, 71, 69, 60, 50, 49, 59

Compute the power spectrum of these data and assess statistical significance of spectral peaks. Do this as a pencil and paper exercise, showing all work in your calculations.

Follow these steps:

1) Subtract out the mean from the data.

2) For each k, compute the coefficients A_k, B_k, and C_k using a discrete Fourier transform.

3) Compute and plot the power spectrum a $\Phi(k)$ as a function of k or ω.

4) Superimpose the theoretical red noise spectrum and assess statistical significance of spectral peaks at the 95% confidence level. Make sure to explain the procedure you use to compute both. Note you are just considering the complete line spectrum in this case, so each spectral line only has 2 DOF.

5) Verify that the results match those of whatever computational tools you will use for the rest of this assignment (e.g. your own programming of the discrete Fourier transform or canned FFT).

Discussion: Given the data, what would be the dominant frequency or frequencies you would expect to appear a priori in the power spectrum? Is this verified through your analysis?

Part II

Calculate and plot the power spectrum for the both the hourly and by minute direct normal solar radiation, using the analysis constraints described below. In all plots, superimpose the estimate of $\Phi(k)$ with corresponding theoretical red noise spectrum. Assess the statistical significance of any spectral peaks at the 95% confidence level. In the following, N corresponds to the total number of time steps. You should have two plots for each case (one for each time series).

Case a) *Boxcar window, no spectral averaging or smoothing. (the "don't know any better" case)*
Lowest resolved frequency: 2π radians per N.
Bartlett (or boxcar) window

No smoothing of the spectra

Case b) *Boxcar window, spectral smoothing*
Lowest resolved frequency: 2π radians per N.
Bartlett (or boxcar) window
Smooth spectra using a 5-point running mean filter (i.e. estimate for $\Phi(k)$ is found by averaging all values of $\Phi(k)$ of the original spectrum between points $\Phi(k-2)$ and $\Phi(k+2)$

Case c) *Boxcar window, spectral averaging*
Lowest resolved frequency: 2π radians per N/10.
Average spectral calculated for 10 subsets of the data, corresponding to 1:100,101:200…etc.

Case d) *Hanning window, spectral averaging with overlap*
Lowest resolved frequency: 2π radians per N/5.
Use an overlap corresponding to half the length of each subset of the data. This means calculating spectral for 9 subsets of the data corresponding to elements 1:200; 101:300;…801:1000.

Discussion: Outline the step-by-step procedures you performed to generate these spectral analyses. What are the relative advantages and disadvantages of the various techniques to smooth, average, or data taper (using a window) on the spectral analysis of these data? Where are the statistically significant peaks found, if any, and do these make physical sense? Are there any differences in statistically significant peaks with higher time resolution? Does significance change depending on the assumptions for the four analysis cases a-d? Why might this be the case?

Part III

Use whatever analysis constraints you deem appropriate to determine the dominant timescale(s) of variability of the climate indices used in Assignment #4 (MEI, PDO index, AO index). Consider all the data available, so the time resolution is one month. Do not normalize the data in any way. Explain how the significance of any spectral peaks is assessed. *A note of caution: just because phenomena are called oscillations does not necessarily mean they will pass a significance test for any way you may construct the spectral analysis!*

Discussion: Compare your results with any published studies or any information provided on the CDC website or similar climate websites. Do your results confirm to the a priori expectations of the dominant timescales of these indices from documentation on the web or in the literature?

Notes and Hints (Professor Thompson's and my own)

The power spectrum is estimated from C_k^2.

I suggest normalize so that the total variance (area under the curve) is one. When testing the significance of a peak, make sure the area under the red noise spectrum = area under spectrum for original data.

Degrees of freedom for the spectrum are calculated as the number of data points divided by the number of independent spectral estimates.

$\omega=2\pi k/N$ where k is the wavenumber. You may plot the x-axis on your plots as ω, k, in terms of numbers of time steps, or physical units of time (e.g. months or years). If plotting as a function of time steps, the first value of the x-axis corresponds to the total number of time steps being analyzed (*k=1,* $\omega=2\pi/N$) and the last value corresponds to the inverse of the Nyquist frequency (two times the time step, $k=N/2$, $\omega=\pi$).

If using a canned FFT routine, many of them give the A_k's and B_ks for the discrete Fourier transform in complex notation (e.g. $A + Bi$). So calculate the complex conjugate to find the corresponding value of C_k^2 (i.e. $C^2 = (A + Bi)(A – Bi)$). *Make sure you know exactly what your black box is doing, in any case, if you choose to go this route!*

Use the theoretical red noise spectrum, derived from the e-folding timescale, as discussed in class and in the Hartmann notes. Remember that statistical significance is calculated as a function of the red noise using a Chi-squared or F statistic.

Assignment due date: Week of finals when term projects due.