ATOC7500 – Application Lab #4
Spectral Analysis of Timeseries
in class Monday October 19 and Wednesday October 21

ASK IF YOU HAVE QUESTIONS ☺

Notebook #1 – Spectral analysis of hourly surface air temperatures from Fort Collins, Colorado at Christman Field
ATOC7500_applicationlab4_fft_christman.ipynb

LEARNING GOALS:
1) Complete a spectral analysis using two different functions in Python (direct FFT from numpy and using scipy which has more options). Describe the results including an interpretation of the spectral peaks and an assessment of their statistical significance.
2) Contrast applying a Boxcar and a Hanning Window when calculating the power spectra. What are the advantages/disadvantages of these two window types? What are the implications for the resulting power spectra?

DATA and UNDERLYING SCIENCE:
In this notebook, you analyze two years (January 1, 2013 thorugh December 31, 2014) of hourly surface temperature observations from Christman Field in Fort Collins, Colorado. Missing data have been already treated. The data are in .csv format and are called Christman_data_nomissing.csv.

Questions to guide your analysis of Notebook #1:

1) Look at your data. What are the autocorrelation and e-folding time of your data? What spectral peaks do you expect to find in your analysis and how much power do you think they will have?

The AR1 is equal to 0.99 and the e-folding time is 100.9 hours. I expect to find spectral peaks at daily and yearly frequencies with power that combined explains about half of the variance.

2) Calculate the power spectra using the Numpy method, which assumes a Boxcar window that is the length of your entire dataset. Graph the power spectra, the red noise fit to the data, and the 99% confidence interval. What statistically significant spectral peaks did you find? What do they represent? How did you assess the statistical significance (what is the null hypothesis that you are trying to reject)? Compare back to Barnes and Hartman notes to make sure all of the equations and functions in the notebook are working as you expect them too.

I find statistically significant peaks at frequencies which correspond to 365, 1, and 0.5 days. These peaks correspond to seasonal, daily, and diurnal variability respectively. You assess the statistical significance by testing against the null hypothesis that the peak is red-noise associated with a red-noise fit to the time series.

3) Calculate the power spectra using the scipy method. Check that you get the same result as you got using the Numpy method. Next – compare the power spectra obtained using both a Boxcar window and a Hanning window. Assume a window length that is the entire length of the dataset. Do you get the same statistically significant peaks when applying the Hanning window and the Boxcar window? How do they differ? Can you explain why?

Indeed, both methods produce the same result. In general, the Hanning window produces the same significant spectral peaks however does lead to spectral smoothing when compared to the Boxcar window. This is because the central lobe of the Hanning window widens in order to remove the effects of the high frequency power needed to resolve the edge of the boxcar window.

4) If time – take a look at other surface meteorological variables in the dataset. Do you obtain similar spectral peaks?

For relative humidity, I only find statistically significant spectral peaks at 24- and 12-hour periods. Interesting! I guess any yearly cycle is specific humidity is negated by air temperature cycles in relative humidity.

<u>Notebook #2 – FFT analysis using Dome-C Ice Core Data</u>
ATOC7500_applicationlab4_fft_EPICA.ipynb

LEARNING GOALS:
1) Calculate power spectra of a dataset available on a non-uniform temporal grid. Describe the results including an interpretation of the spectral peaks and an assessment of their statistical significance.
2) Contrast applying a Boxcar and a Hanning Window when calculating the power spectra. What are the advantages/disadvantages of these two window types? What are the implications for the resulting power spectra?
3) Apply a Hanning Window with various window lengths - What are the advantages/disadvantages of changing the window length and the implications for the resulting power spectra in terms of their statistical significance and temporal precision?
4) Apply a Hanning Window with various window lengths and use Welch's method (Welch's Overlapping Segment Analysis, WOSA). How does WOSA change the results and why?

DATA and UNDERLYING SCIENCE:
In this notebook, you will perform a power spectral analysis of the temperature record from the Dome-C Ice Core, taken at 75 South and 123 East (Jouzel et al. 2007). The temperature data go back ~800,000 years before present. They are unevenly spaced in time. The data are available on-line here, courtesy of the NOAA Paleoclimatology Program and World Data Center for Paleoclimatology:
ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/antarctica/epica_domec/edc3deuttemp2007.txt More information on the data is available at:
https://www.ncdc.noaa.gov/paleo-search/study/6080

Questions to guide your analysis of Notebook #2:

1) <u>Look at your data and pre-process for FFT analysis:</u> Power spectra analysis assumes that input data are on an evenly spaced grid. The Dome-C temperature data are not uniformly sampled in time. Regrid the Dome-C temperature data to a uniform temporal grid in time. Plot the data before and after re-gridding to make sure the re-gridding worked as expected.

Yes, both time series appear generally similar, although the high frequency climate information from the last ~200,000 years is now lost ☹ So in essence yes, both time series make sense, but we are now deliberately removing some information!

2) <u>Signal and Noise:</u> What is the autocorrelation and e-folding time of your data? What spectral peaks do you expect to find in your analysis and how much power do you think they will have? *Hint: Think back to the Petit 1999 Vostok ice core dataset discussed in class.*

The lag-1 autocorrelation is 0.96 and the e-folding time is 25 years. I expect to find significant spectral peaks at frequencies which correspond to the Milankovitch cycles, which occur at roughly 100,000, 41,000, and 24,000 year cycles.

3) Use Boxcar Window to calculate power spectra: Calculate the power spectra using the Numpy method, which assumes a Boxcar window that is the length of your entire dataset. Graph the power spectrum, the red noise fit to the data, and the 99% confidence interval. What statistically significant spectral peaks did you find? What do they represent?

Significant spectral peaks exist at frequencies of 0.01, 0.025, and 0.044 millennia^{-1} which correspond to the Milankovitch cycles.

4) Compare Boxcar Window vs. Hanning Window: Calculate the power spectra using the SciPy method. Compare the results obtained using a Boxcar window that is the length of your entire dataset to those obtained using a Hanning window that is the length of your entire dataset. Graph the power spectrum, the red noise fit to the data, and the 99% confidence interval. What statistically significant spectral peaks did you find? What do they represent? What are the differences between the results obtained using the Boxcar window and the Hanning window? Is the intuition that you gained by looking at Fort Collins temperatures the same as what you are seeing here with Dome-C temperature records? Why or Why not?

Both windows produce the same three statistically significant spectral peaks at frequencies of 0.01, 0.025, and 0.044 millennia^{-1}. These three spectral peaks correspond to the Milankovitch cycles in orbital eccentricity, obliquity, and precession.

As was the case in the Fort Collins temperature time series, the peak power associated with the boxcar window is greater than that of the Hanning window which is instead more diffused. Interestingly the peak powers in both windows do not always align at the same frequency. Something interesting to consider is that because the Hanning window produces more diffused spectra, it could be more difficult to reject the null-hypothesis that the spectral peak is simply red-noise memory.

5) Hanning Window with different window lengths: Using the SciPy method, compare the power spectra obtained using Hanning window with different window lengths. Graph the power spectra, the red noise fit to the data, and the 99% confidence interval. Did you find any statistically significant spectral peaks? How does decreasing the window length affect the temporal precision of the spectral peaks and their statistical significance? Did you find the classic tradeoff between 1) high spectral/temporal resolution but low-quality statistics, and 2) high quality statistics but low spectral/temporal resolution?

This is quite interesting! By reducing the window length, we no longer find a significant spectral peak corresponding to a cycle of 100,000 years. Decreasing the spectral resolution diffuses the spectral peaks, which is similar to what we found when comparing the Boxcar and Hanning windows. The 100,000 year cycle is right at the Nyquist frequency and therefore cannot be expected to be properly resolved.

Indeed I do see the tradeoff between spectral accuracy and statistical significance!

5) Add WOSA (Welch Overlapping Segment Averaging): Having found what you think is a good balance between precision in the identification of the spectral peaks and statistical significance – Try applying WOSA (Welch Overlapping Segment Averaging) in addition to using the Hanning Window with different window lengths. How does this change your results?

When using WOSA the number of degrees of freedom is the same as using just one window of length 200,000 years (so the 99% significance level remains the same), however by adding additional information we retain information at the edges of the window. This fact allows us to detect a significant spectral peak at a frequency corresponding to the 100,000 year Milankovitch cycle.