<u>ATOC5860 – Application Lab #4</u>
<u>Spectral Analysis of Timeseries</u>
<u>in class March 10 and March 15</u>

ASK IF YOU HAVE QUESTIONS ☺

<u>Notebook #1 – Spectral analysis of hourly surface air temperatures from Fort Collins, Colorado at Christman Field</u>
ATOC5860_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 through 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 autocorrelation is 0.99 and the e-folding time is 100 hours, or about 4 days. We expect to find a diurnal and annual cycle, where the annual cycle exhibits a stronger signal due to a greater seasonal temperature variation.

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.

We find statistically significant spectral peaks at annual, daily, and semi-daily frequency. The annual cycle represents the seasonal cycle, the daily cycle represents daytime vs. nighttime, and I'm not sure what the 12-hour cycle stands for. The null *hypothesis is that the peak signals do not exceed that of the red noise fit. Each peak is significant at a 99% confidence level.*

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?

We do obtain the same power spectra with the scipy method. We do obtain the same statistically significant peaks with both the Hanning and Boxcar windows. However, the Hanning window produces wider peaks due to the tapered ends of the weighting function.

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

Relative humidity lacks an annual spectral peak but also exhibits diurnal and semi-diurnal spectral peaks. Wind speed also exhibits annual, diurnal, and semi-diurnal spectral peaks but the annual peak is not as significant as the temperature peaks.

Question: Are you seeing power at 12-hour frequencies when looking at temperature? Maybe it is atmospheric tides? Or is it some kind of spectral ringing artifact? Unsolved mysteries of ATOC7500 Objective Data Analysis…

<u>Notebook #2 – FFT analysis using Dome-C Ice Core Data</u>
ATOC5860_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.

We can regrid the data by applying np.interp to an np.linspace object. This allows us to project our irregular values onto an evenly spaced temporal array where each timestep is one millennium.

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.*

I expect to find three spectral peaks which correspond to the Milankovich cycle: procession at 19,000 to 24,000 years, eccentricity at 100,000 years, and obliquity/tilt at 41,000 years. I assume that eccentricity will yield the highest power because it has the longest frequency.

3) <u>Use Boxcar Window to calculate power spectra:</u> 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?

I found three statistically significant spectral peaks: one at about 100,000 years, one at about 40,000 years, and one at about 24,000 years. These three peaks match with the expected phases of the Milankovich cycle mentioned in question 2). The absolute most powerful peak is the eccentricity peak, however the obliquity peak is the most powerful relative to the red noise curve.

4) <u>Compare Boxcar Window vs. Hanning Window</u>: 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?

I find the same three statistically significant spectral peaks regardless of method as I found in question 4). The Hanning window leads to broader, lower-power spectral peaks, to the point where procession only just appears as statistically significant.

5) <u>Hanning Window with different window lengths:</u> 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?

Yes, although the shorter window length of 200 millennia did not produce a statistically significant peak on the order of eccentricity. For a total time window of 400,000 years, the cycle could theoretically only contain 3-4 eccentricity peaks, so only a slight variance in the eccentricity cycle "dilutes" the signal across spectral frequencies. At the same time, increasing the window to 800 millennia recognizes the three spectral peaks but the procession and obliquity peaks appear sharper. That is, the variance in procession and obliquity is blurred in the longer timeframe. Decreasing the window length does in fact produce lower quality statistics but higher temporal resolution.

5) <u>Add WOSA (Welch Overlapping Segment Averaging):</u> 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?

WOSA over Hanning windows allows for the identification of all three spectral peaks at a window length of 200 millennia. Shrinking the window length too far introduces the same issue as in problem 4). .Interestingly, the number of independent samples following the Leith method is only N=16.*