**ATOC5860 – Application Lab #4**

**Spectral Analysis of Timeseries**

**in class March 10 and March 15**

ASK IF YOU HAVE QUESTIONS ☺

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

**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?

There looks to be a lot more autocorrelation during the winter months. Based on this plot and what I know about temperature, I expect to see a frequency that shows up at 24 hours and 1 year. A lag-1 autocorrelation of 0.99 was calculated. We found an e-folding time of 100.92 hours.

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.

There are three significant spectral peaks. One is at approximately 0.000110 hours, which corresponds to 1 year. There is another peak at a 24 hour period, and another at a 12 hour period. To assess the spectral significance, we compare the height of the peak to a red-noise power spectrum. Our null hypothesis is that the observed frequency is purely due to memory in the data, and not due to an underlying cycle. The only way I could explain this is a certain nonlinear process, such as clouds, would speed up/slow down temperature change depending on the time of day.

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?

The scipy method seems to give the same three frequencies. The peaks are all in the same spot and statistically significant. The Hanning peaks are slightly wider than the Boxcar peaks, but not to the point where it makes a peak insignificant. The Hanning window “smoothens” the peaks, so this makes sense.

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

I am very confused by the 12-hour frequency!

**Notebook #2 – FFT analysis using Dome-C Ice Core Data**

**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. Look at your data and pre-process for FFT analysis: 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.

Chart

Description automatically generated

Chart

Description automatically generated

It worked! The data are now evenly spaced.

1. Signal and Noise: 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.*

Our lag-1 autocorrelation is 0.96 days, with an e-folding time of 25 kyr. I expect to see peaks at all three major Milankovitch Cycles. They would give me frequencies of 0.01, 0.025, and 0.043 kyr^(-1). These frequencies correspond to eccentricity, obliquity, and precession respectively.

1. 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?

I found all three peaks that I predicted. They are all above the red noise fit at 99% confidence.

**Chart, line chart

Description automatically generated**

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?

The same three peaks are found using SciPy. Again, the Hanning peaks are wider than the Boxcar peaks, but all three orbital peaks remain significant. Again, these represent the three Milankovitch Cycles. I am interested in the slight shift in location of the Hanning window peak at 100 kyr, it appears to be picking up the small peak at around 80 kyr.

Chart, line chart

Description automatically generated

1. 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?

Altering the window length changes the resolution of the peaks given. A shorter window gives us high-quality statistics with a greater requirement for significant. A longer peak gives us better resolution, but we can be less confident in the peaks provided.

1. 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?

WOSA gives a very weak peak at 100 kyr, but only has a significant peak for obliquity.