**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?

The lag-1 autocorrelation is .99 and the e-folding time-scale is 100.92 hours. We expect to see power at the annual and diurnal cycle.

1. 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 found spectral peaks corresponding to 365 days and 1 day, as we expected. Additionally, we saw anoter stastically significant peak right by the 1 day cycle. This is probably due to the broadening of the peaks due to the discreet nature of the fft. Additionally we saw something at ~12hr mark. Need to assess if this a numerical artifact. (or something about the tides???).

To assess statistical significance, we construct a 99% CI using an f-stat. for the PDS of red noise with the null hypothesis that our signal’s PSD comes from red noise. If our value falls outside of this CI, we can reject the null hypothesis. We can say with 99% confidence that this frequency is not just from red noise.

1. 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?

When using np vs scipy, we do get the same statistically significant peaks, however under the threshold we see some variation in the two outputs. This is surprising as scipy *should* be build around numpy. There may be differences in implementation for this specific function.

When we use different windows, we get the same peak location but different shapes; This makes sense due to the convolution in the frequency domain “smoothing” out the PSD.

*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…

**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.

2) 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.*

The autocorrelation is .96 and the e-foilidng time scale is 25k years.

According to my glacially-inclined classmates, we should expect to see the Milankovitch cycles: 100,000, 40,000, ~20,000 years

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?

Chart, line chart

Description automatically generated

We see peaks corresponding to 100328, 40131, 23607, and 22932 years. We also could expect to see a value corresponding to 400,000 years, however the range of our data is not long enough to this peak with statistical significance.

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?

Chart, line chart

Description automatically generated

With both windows, we are able to see the same spectral peaks, however the hanning window (as expected) has broader and shorter peaks. Both methods still fall above the CI from the f-test. This agrees with our analysis of the temperature.

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?

Graphical user interface, chart

Description automatically generated

We can see that when using a shorter window, we lose the significant peak from the slowest frequency (~100k years). We are not able to fit a full cycle of this frequency into our window. As we decrease the window length, we gain more power, but lose frequency specificity. In contrast, as we get more spectral resolution our peaks fall closer to the CI for red noise.

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 we use WOSA and hanning window, get all of the peaks back that we expected. We notice that the power is the highest in this method and the peaks are just above the red noise CI.**