**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?

Temperature lag-1 autocorrelation = 0.99 and Te = 100.92 hours. It needs 100 hours to decrease by 60% in it’s memory. We should see a yearly cycle and a daily period and the yearly cycle is the one with more power because it would have a higher amplitude compared to the daily cycle. 12 hour peaks

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.

##### FOUND IT - spectral peak exceeds red noise ####

exceeds by... 0.516

at frequency.... 0.0001

which in days is... 365.0

##### FOUND IT - spectral peak exceeds red noise ####

exceeds by... 0.115

at frequency.... 0.043

which in days is... 1.0

##### FOUND IT - spectral peak exceeds red noise ####

exceeds by... 0.001

at frequency.... 0.042

which in days is... 0.99

##### FOUND IT - spectral peak exceeds red noise ####

exceeds by... 0.004

at frequency.... 0.083

which in days is... 0.5

XXXX FOR REFERENCE XXXX

12-hourly in frequency: 0.08

24-hourly/daily in frequency: 0.04

yearly in frequency: 0.00011

What is the 12 hour peak? Atmospheric tidal – semi diurnal tide in the wind field.

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?

Boxcar gives higher peaks and therefore more statistically significant.

*Chart, line chart

Description automatically generated4) If time – take a look at other surface meteorological variables in the dataset. Do you obtain similar spectral peaks?*

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

**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) 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.

Before re-gridding:

Chart

Description automatically generated

After re-gridding:

**Chart

Description automatically generated**

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.*

Temp lag-1 autocorrelation = 0.96 and Te = 25

25 \* 1000 where one timestep is a millenia which means that the memory will decrease over 1/e. So the periods might be ice ages every hundred thousand year.

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?

##### FOUND IT - spectral peak exceeds red noise ####

exceeds by... 0.0941063100164142

at frequency.... 0.01

which in years is approximately... 100328.0

**eccentricity**

##### FOUND IT - spectral peak exceeds red noise ####

exceeds by... 0.10688859416323286

at frequency.... 0.025

which in years is approximately... 40131.0

**obliquity: 41,000 years**

##### FOUND IT - spectral peak exceeds red noise ####

exceeds by... 0.01891507042813571

at frequency.... 0.0425

which in years is approximately... 23607.0

**procession 25,700 years**

##### FOUND IT - spectral peak exceeds red noise ####

exceeds by... 0.006414730279733135

at frequency.... 0.043750000000000004

which in years is approximately... 22932.0

**procession**

**earths orbital cycles (eccentricity: shape of the earth orbit) (obliquity is tilt) (procession…) are the first peak, which lead to glacial and inter glacial periods which is the next peak.**

**When we look at it we see that the last two are under the same peak:**

**Chart, line chart

Description automatically generated**

##### FOUND IT - spectral peak exceeds red noise ####

exceeds by... 0.0941063100164142

at frequency.... 0.01

which in years is approximately... 100328.0

##### FOUND IT - spectral peak exceeds red noise ####

exceeds by... 0.10688859416323286

at frequency.... 0.025

which in years is approximately... 40131.0

##### FOUND IT - spectral peak exceeds red noise ####

exceeds by... 0.01891507042813571

at frequency.... 0.0425

which in years is approximately... 23607.0

##### FOUND IT - spectral peak exceeds red noise ####

exceeds by... 0.006414730279733135

at frequency.... 0.043750000000000004

which in years is approximately... 22932.0

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? Yes? Why or Why not?

Chart, line chart

Description automatically generated

We found peaks at ~frequencies of 0.01, 0.025, and 0.043.

Boxcar can produce best most accurate peaks and frequencies but can produce fake peaks while the Hanning window typically projects low and wider frequencies. Therefore, you can use Boxcar to find the peaks and % confidence and then check if they are real using the Hanning window.

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?

One at 0.024 and 0.04 and 0.045

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?

If we chunk too small we gotta chuck it because the frequencies we would have found in the larger window will not be found and will start influencing the smaller window frequencies:

When you reduce the window size you reduce the number of frequencies you can find. Because the windows chunk the data: if you have 1k, full window = n number of estimates. If window is 100, then you take the first 100 and get estimate, then the second 100 etc. you get less estimate. “### Take-home message for 100,000ish year peak. If you reduce your window length too much -- your data chunks

### are too short to see the low frequency oscillations.”

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?

At first glance, the WOSA and Hanning together make a clean, three peaks at frequencies of 0.01 0.025 and 0.045. The WOSA uses overlaps of the tapered windows but doesn’t discount the data, that is still important, at the edges. The WOSA applies windows with 50% overlapping windows thus gaining two degrees of freedom. The reason each chunk is not double counting is because you are able to take advantages of correlations in the overlapping spaces that you wouldn’t have used otherwise.

Overall things are more significant.