**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?

*Temperature lag-1 autocorrelation = 0.99 and Te = 100.92 hours. I would expect to find big peaks around the seasonal and diurnal cycle. Diurnal probably more power than seasonal since the seasonal cycle has a lower frequency.*

**Chart, line chart

Description automatically generated**

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... 0.5163757684804352

at frequency.... 0.00011415525114155251

which in days is... 365.0

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

exceeds by... 0.11520792476140507

at frequency.... 0.041666666666666664

which in days is... 1.0

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

exceeds by... 0.0010714417330250896

at frequency.... 0.04178082191780822

which in days is... 0.9972677595628414

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

exceeds by... 0.0038905346383531344

at frequency.... 0.08333333333333333

which in days is... 0.5

XXXX FOR REFERENCE XXXX

12-hourly in frequency: 0.083

24-hourly/daily in frequency: 0.042

yearly in frequency: 0.00011

**A screenshot of a computer

Description automatically generated with medium confidence**

**Graphical user interface, application

Description automatically generated**

*We found three strong power spectra associated with the annual, daily, and diurnal cycle. The null hypothesis would be that none of the power spectra seen here exceed the values compared to a red noise dataset with 99% certainty. This red noise dataset is based on the basic statistics of our dataset.*

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?

**A screenshot of a computer

Description automatically generated with medium confidence**

**A screenshot of a computer

Description automatically generated with low confidence**

**Chart

Description automatically generated**

**Chart, line chart

Description automatically generated**

*Overall, we do get the same peaks between the two methods. The math between the two methods is the same, they just use different python functions to accomplish it.*

*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.

**Chart

Description automatically generated**

**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.*

*I expect the main peaks are due to the Milankovitch cycles, with everything else being due to oscillations on much shorter time scales.*

*Temp lag-1 autocorrelation = 0.96 and Te = 25.0*

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**

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

*exceeds by... 0.0941063100164142*

*at frequency.... 0.01*

*which in years is approximately... 100328*

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

*exceeds by... 0.10688859416323289*

*at frequency.... 0.025*

*which in years is approximately... 40131*

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

*exceeds by... 0.01891507042813571*

*at frequency.... 0.0425*

*which in years is approximately... 23607*

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

*exceeds by... 0.006414730279733131*

*at frequency.... 0.043750000000000004*

*which in years is approximately... 22932*

*It appears that the first three peaks do in fact coincide with the Milankovitch cycles. The 100k cycle is due to eccentricity, the 40k cycle is due to changes in obliquity (Earth’s tilt). The last two (likely combined) is the axial precession (Earth’s wobble).*

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

*It seems we got the same peaks, but the Hanning window appears to combine the two 20ish-k cycles that the Boxcar method had left out. This is probably better (Hanning) as these two 20ish-k cycles are the same phenomena.*

*Overall, I would say what we are seeing here is more pronounced than what we saw in Fort Collins. Here we see the phenomena are more pronounced, and the effect of using a separate 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? 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, application, table, Excel

Description automatically generated

*Decreasing the window has a huge effect, as it completely cuts out the 20k-ish peak (not statistically significant). Adding length to the timeseries it comes back in. Overall, it is probably better to have lower temporal resolution with higher quality statistics, especially with data such as ice cores.*

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?

**Chart, line chart

Description automatically generated**

*Using the default configuration in the notebook, it seems that WOSA helps with this resolution/statistics balance. We do get that 20k-ish statistically significant peak, but barely. The rest of the peaks seem to be captured as well.*