**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?

Autocorrelation: 0.99

E-folding timescale: 100.92 hours

We would expect to find spectral peaks at 1 day (24 hours) and 365 days. We hypothesize that the annual cycle (seasonal cycle) will have a larger peak amplitude compared to the daily (diurnal cycle).

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.

**A screenshot of a computer

Description automatically generated with medium confidence**

The peak with the largest amplitude (explains the most variance) is the annual cycle with a peak period of 365 days. The diurnal cycle is also shown as a peak at 1 day but with less amplitude than the annual cycle. There are a few other significant spectral peaks (above 99% confidence level) at 0.99 days and 0.5 days but they explain a very small percent of variance in the data. We assessed the statistical significance by comparing the spectral peak to the red noise spectrum fitted to the data at the 99% confidence level. The null hypothesis that we are trying to reject is that the peak is only due to random fluctuations of a red-noise spectrum.

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

Chart

Description automatically generated

When applying the Hanning window and the Boxcar window we do see the same peaks however the amplitude it slightly different. For the Hanning window, the amplitude is slightly smaller and the peak frequency has a broader window compared to the boxcar window. However, the boxcar window does seem to have a side lobe on the right side that is at the 99% significance level. This highlights the advantage of using a Hanning tapered window instead of a boxcar window.

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

For wind speed the main spectral peaks still occur at 365 days and 1 day but the annual cycle is now much smaller in amplitude compared to the daily/diurnal cycle.

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.*

Autocorrelation: 0.96

E-folding timescale: 25,000 years

We expect to see peaks around the timescales of the Milankovitch cycles. Obliquity has the biggest impact on climate which would be a peak at 41,000 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

* 1. peak frequency: 100,328 years

0.025 peak frequency: 40,131 years

0.0425 - 0.4375 peak frequency: 23,607- 22,932 years

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

For the Hanning window, the spectral peaks are broader, but we still see peaks at the same frequencies as for the Boxcar window. I think in general, the difference between the Hanning and Boxcar windows is the same for both datasets. Although, the temperature data showed very clear peaks using both the Hanning and Boxcar window, however, the Dome-C core temperature results show much broader peaks in the spectrum for 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? 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

For a smaller window length, the spectral peaks become less clear and have less power, especially for the peaks at lower frequencies. This is likely because with a smaller window, it is more difficult to pick up the lower frequencies.

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**

**Chart, line chart

Description automatically generated**

Using WOSA helps the peaks still appear even with a smaller window length. Although the red noise spectrum is much closer to the peaks, the significant peaks still are clearly visible and align with what we say for the full dataset.