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?

Autocorrelation = 0.99, e-folding time =100.92 hours
I expect it to have a yearly cycle and a daily cycle. The power of the yearly cycle is larger than the power of the daily cycle since temperature range through out a year is larger than the temperature range of a day and power is related to amplitude.

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.

I found peaks at 365 days, 1 day and 12 hours, which represents to seasonal cycle, diurnal cycle and I don't know what's the half day cycle. They go much beyond the red-noise power so it's significant.

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 peak of the boxcar is higher and more concentrate, but it may lead to sub-peak near the main peak. The Hanning window is lower and spreads out a little bit, but one peak remains as just one peak.

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

<u>Notebook #2 – FFT analysis using Dome-C Ice Core Data</u>
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.

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 time is 25 time step which is 25*1003=25075 years. I think the peaks should be related to earth's orbit change, the eccentricity cycle, the obliquity cycle and the precession cycle.

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?

We have peaks at 100328 years, 40131 years, 23607 years and 22932 years. The last two should be the same peak. They are all statistical significant. The first one is Eccentricity cycle (glacial-interglacial cycle) which is about 100,000 years. Second is Obliquity cycle which is about 41000 years. Third is Precession cycle which is 25771 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?

I found the same peaks. The boxcar window is higher and thinner, but Hanning window won't create a fake side-peak. I think they are similar, but the peak seems to shift a little bit between the boxcar and Hanning window which doesn't happen for Fort Collins data and I don't know why.

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?

Decreasing the window length will increase the significance, since we have more pieces of data segment to increase the degree of freedom (the 99% curve will get closer to the red noise solid curve). But each segment is shorter so we may lose long period cycle information and that cycle will be adding up to the higher frequencies to produce fake result. Smaller window length leads to less frequencies it can resolve (lose low frequencies). So for the result, the second and third peaks become more significant, but we lose the first peak for window length = 200.

Longer window length, more accurate, less significant, so I think to be more 保守的,it’s better to use the full data set as the window length.

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?

Use WOSA we don’t lose the first peak with 200 window length. And the second peak is also very significant, but the third peak is less significant compares to the non-wosa one, but not worse than using the whole window length.