<u>ATOC7500 – Application Lab #4</u>
<u>Spectral Analysis of Timeseries</u>
<u>in class Monday October 19 and Wednesday October 21</u>

ASK IF YOU HAVE QUESTIONS ☺

<u>Notebook #1 – Spectral analysis of hourly surface air temperatures from Fort Collins, Colorado at Christman Field</u>
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?

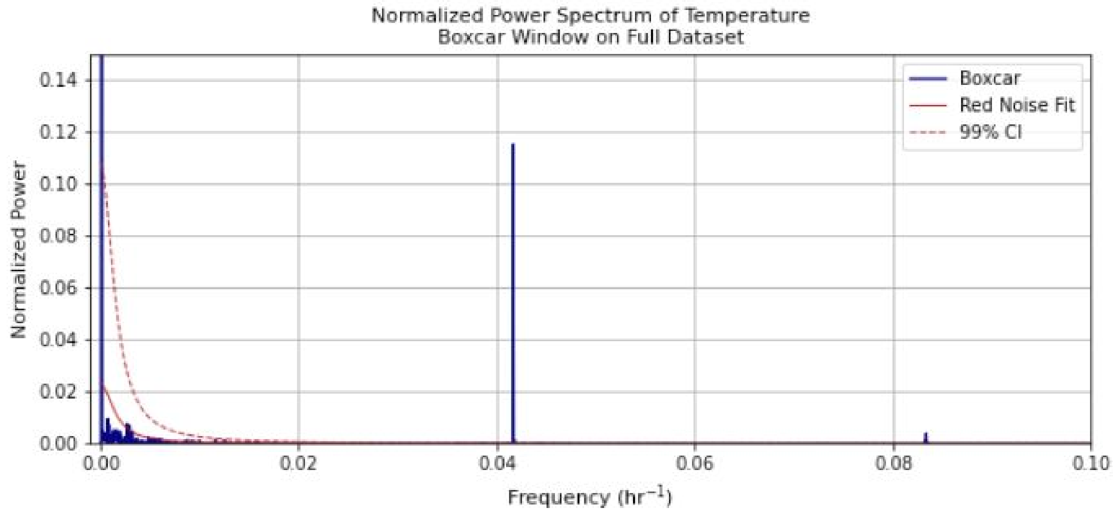
DATA and UNDERLYING SCIENCE:
In this notebook, you analyze two years (January 1, 2013 thorugh December 31, 2014) of hourly surface temperature observations from Christman Field in Fort Collins, Colorado. Missing data have been already treated. The data are in .csv format and are called Christman_data_nomissing.csv.

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?

> The lag1 autocorrelation is 0.99 and the e-folding time is 100.92 hours. Because there is so much memory in the data, I expect there to be two spectral peaks - one for the diurnal cycle (high-frequency peak) and one for the seasonal cycle (low-frequency peak). I expect that the power for the diurnal cycle spectral peak will be greater than the power for the seasonal cycle spectral peak.

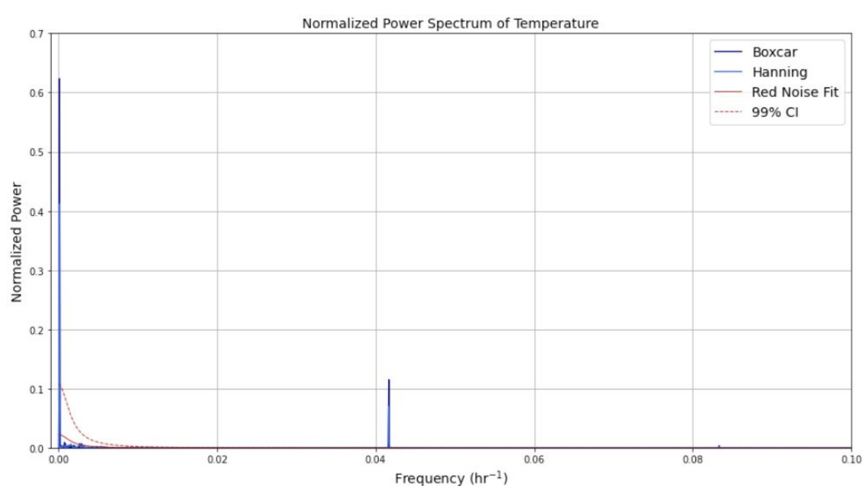
2) Calculate the power spectra using the Numpy method, which assumes a Boxcar window that is the length of your entire dataset. Graph the power spectra, the red noise fit to the data, and the 99% confidence interval. What statistically significant spectral peaks did you find? What do they represent? 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.



There are 4 peaks that exceed the red noise fit at the 99% confidence interval. Once at a frequency of 365 days, once at a frequency of 1 day, another at a frequency of 0.9973 days and a final one at a frequency of 0.5 days. The first peak (365 days) represents the annual cycle. The next two peaks (0.9973 days and 1 day) both represent the diurnal cycle. We assessed the statistical significance by performing an f-test on each peak that exceeded the red noise fit. The null hypothesis was that the data was red noise.

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?

We do get the same power spectra using the scipy method (see below).



The significant peaks are mostly the same using the Boxcar and Hanning windows however, with the Hanning window, you do not get two separate peaks for the diurnal cycle. This is because the Hanning window distributes the power at a wider range of frequencies, which also manifests itself in a lower power value at the spectral peak. With the Boxcar window, you can see that there is significant power at frequencies surrounding the large 1-day spectral peak (0.9973 days). This is likely evidence of side lobes that can occur when using the Boxcar window, adding frequencies that do not occur in the original data.

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?

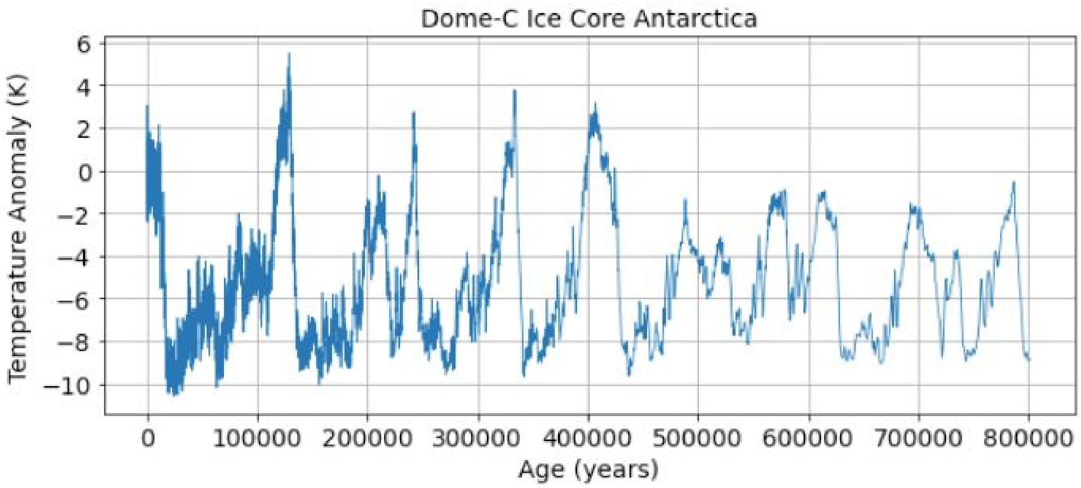
DATA and UNDERLYING SCIENCE:
In this notebook, you will perform a power spectral analysis of the temperature record from the Dome-C Ice Core, taken at 75 South and 123 East (Jouzel et al. 2007). The temperature data go back ~800,000 years before present. They are unevenly spaced in time. The data are 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:

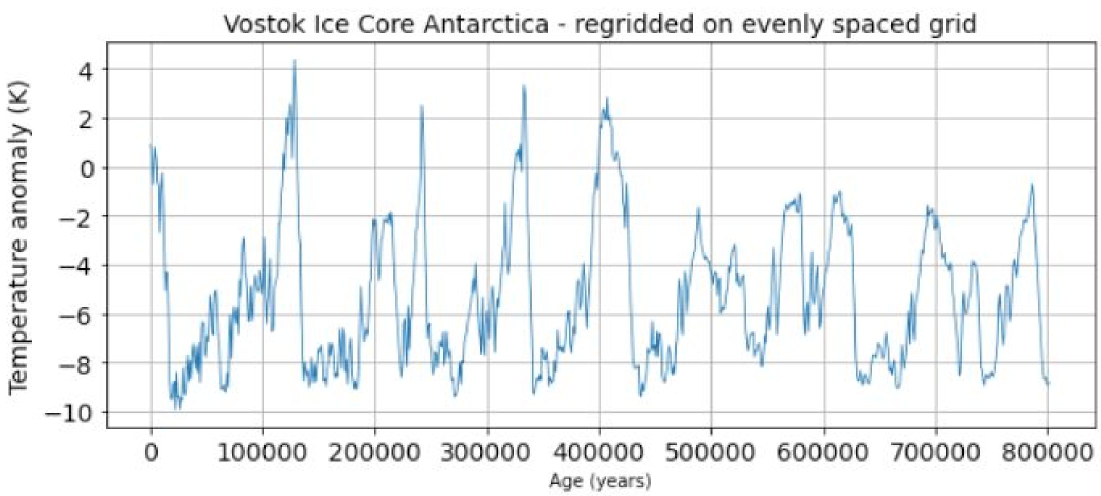
1) <u>Look at your data and pre-process for FFT analysis:</u> Power spectra analysis assumes that input data are on an evenly spaced grid. The Dome-C temperature data are not uniformly sampled in time. Regrid the Dome-C temperature data to a

uniform temporal grid in time. Plot the data before and after re-gridding to make sure the re-gridding worked as expected.

Before regridding:



After regridding:



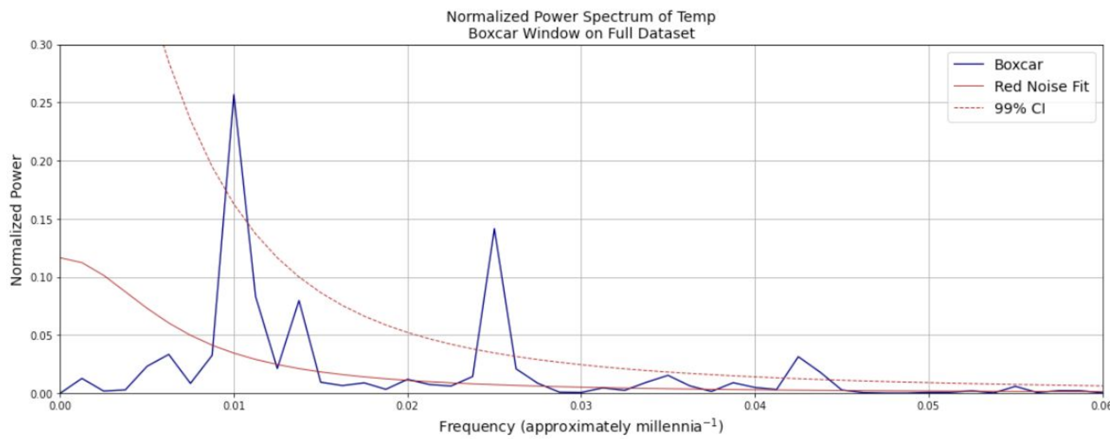
You can see from the plots above that by regridding, we lose some of short-term variation in more recent years (on the left side of the upper plot); however, the long-term patterns remain the same.

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 lag1 autocorrelation is 0.96, indicating that this dataset has a lot of memory. This makes sense because ice cores hold a record of climate patterns which vary at long time scales. The e-folding time is approximately

25,000 years. From visually looking at the data, I expect there to be a spectral peak at approximately 80,000 years with a high power. I also expect to see a spectral peak at approximately 10,000 years with a lower power.

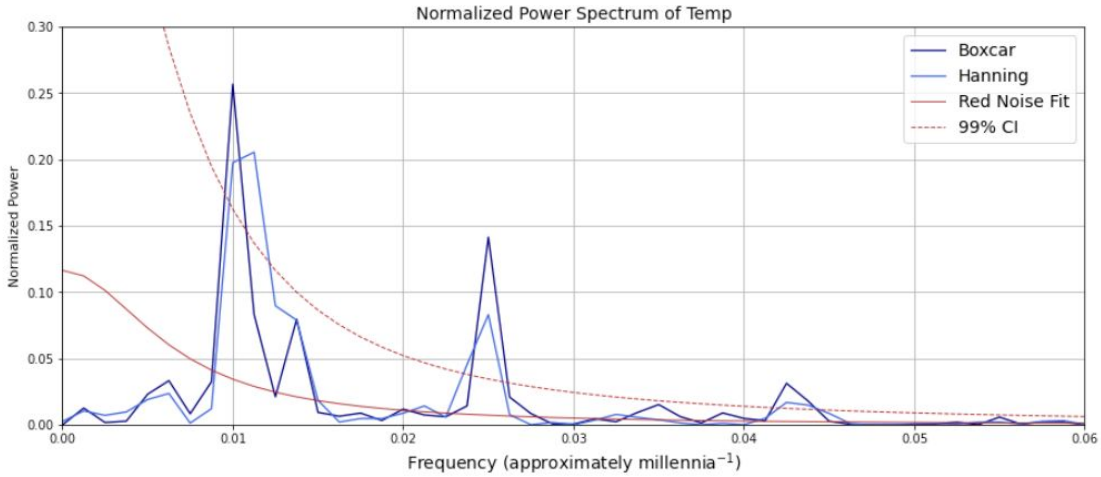
3) Use Boxcar Window to calculate power spectra: Calculate the power spectra using the Numpy method, which assumes a Boxcar window that is the length of your entire dataset. Graph the power spectrum, the red noise fit to the data, and the 99% confidence interval. What statistically significant spectral peaks did you find? What do they represent?



There are 4 statistically significant spectral peaks at the 99% confidence interval. This first is at 100,328 years and represents the Milankovitch cycle of eccentricity. The second is at 40,131 years and represents the Milankovitch cycle of obliquity, or tilt. The third and fourth are at 23,607 and 22,932 years respectively and represent the Milankovitch cycle of precession.

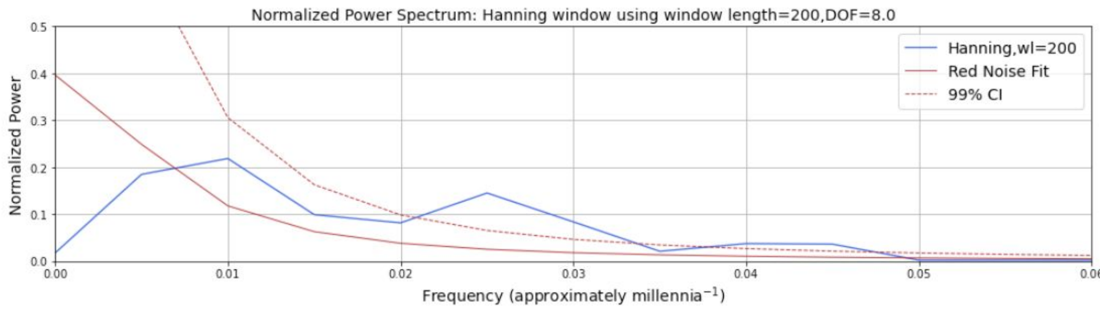
4) Compare Boxcar Window vs. Hanning Window: Calculate the power spectra using the SciPy method. Compare the results obtained using a Boxcar window that is the length of your entire dataset to those obtained using a Hanning window that is the length of your entire dataset. Graph the power spectrum, the 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?

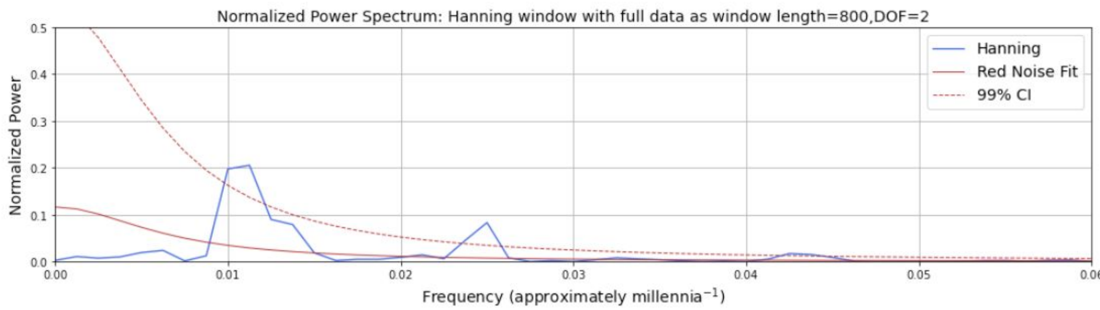
Below is the power spectra using the SciPy method with both a Boxcar window and a Hanning window. We see the same spectral peaks using both methods that we found above using the numpy fft method. We can see that the spectral peaks using the Hanning window have a lower power and are spread out, or smoothed, across a wider range of frequencies. Using the Boxcar window, you can see the results of some 'side lobes', particularly at the lowest frequency 100,000 year peak. However, none of the 'side lobe' peaks are statistically significant. This is the same intuition that we gained from the Fort Collins temperature dataset.



5) <u>Hanning Window with different window lengths:</u> 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?

Below is a comparison of the power spectra obtained using the Hanning window with different window lengths (200 for upper figure and full data for lower figure). You can see that by decreasing the window length, we lose one of the statistically significant spectral peaks: the 100,000 year peak. This is because by decreasing our window size we are no longer able to resolve this low-frequency peak. This provides a good tradeoff between spatial/temporal resolution and statistics. By decreasing the window length we increase the degrees of freedom and thus our results are more statistically robust. However, we lose resolution at the lower frequencies. This method would be appropriate if we care more about the higher frequency signals and don't care if we lose low-frequency resolution.





5) <u>Add WOSA (Welch Overlapping Segment Averaging):</u> 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?

Below is a plot of the power spectra after applying WOSA with a Hanning window length of 200. We can see that we are able to increase the degrees of freedom (8), making the result more statistically robust, while maintaining temporal resolution at the lower frequencies. Using this method we only get 3 statistically significant spectral peaks, each one representing a different Milankovitch cycle.