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?

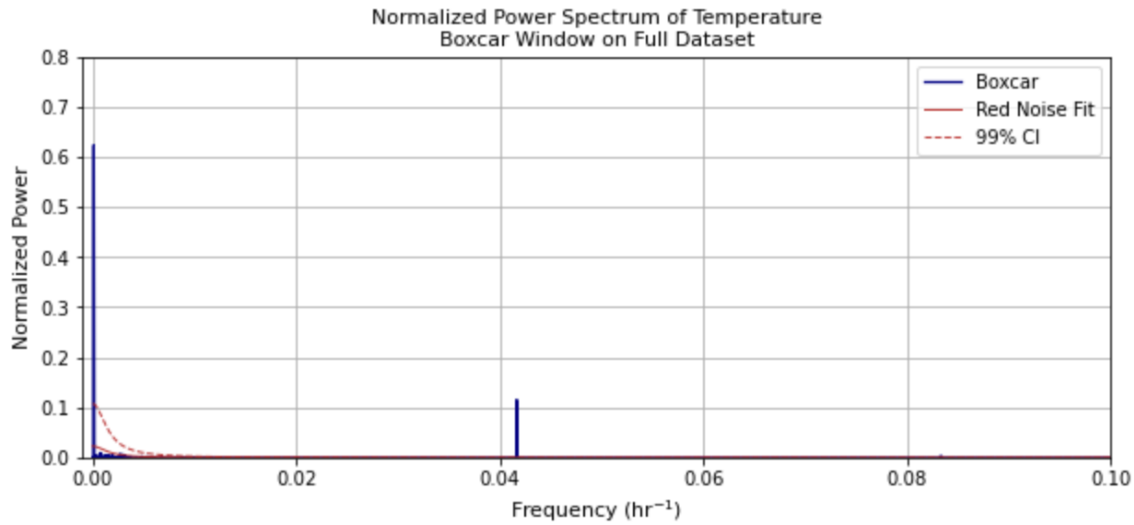
DATA and UNDERLYING SCIENCE:
In this notebook, you analyze two years (January 1, 2013 through December 31, 2014) of hourly surface temperature observations from Christman Field in Fort Collins, Colorado. Missing data have been already treated. The data are in .csv format and are called Christman_data_nomissing.csv.

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?

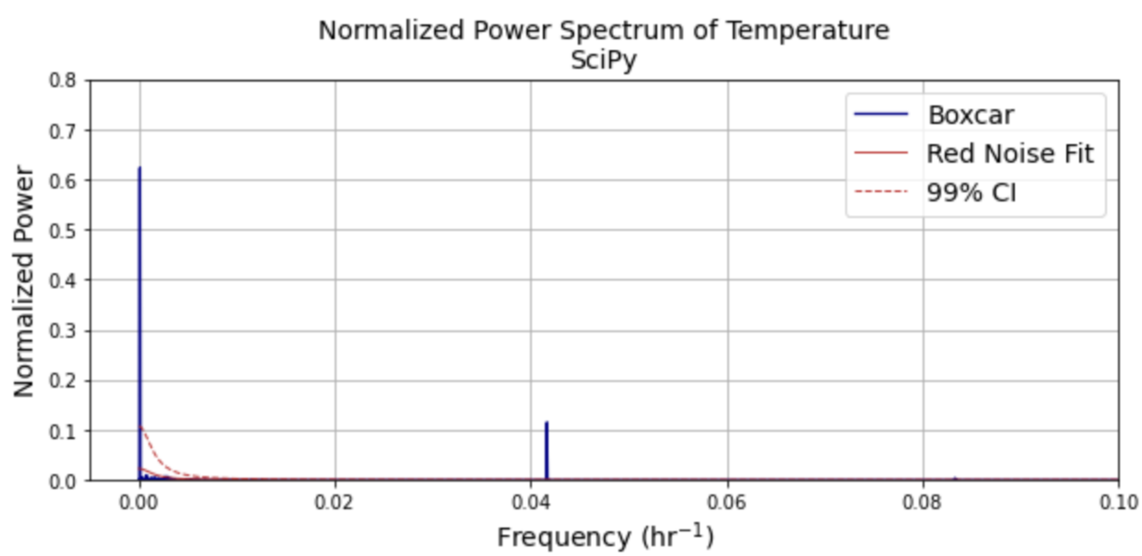
The autocorrelation is 0.99 and the e-folding time is 100.92 hours. We expect to see annual and diurnal.

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.



Four frequencies have statistical significance greater than 99%. Two of these match our expectations (365 days and 1 day) and two do not (0.99 days and 0.5 days). The two that match our expectations and are very statistical significant can be seen in the graph above with frequencies of 0.0001 $hr^{-1}$and 0.04 hr^{-1}. Null hypothesis is that data have the same power spectrum as red noise. We can reject the null hypothesis with 99% significance.

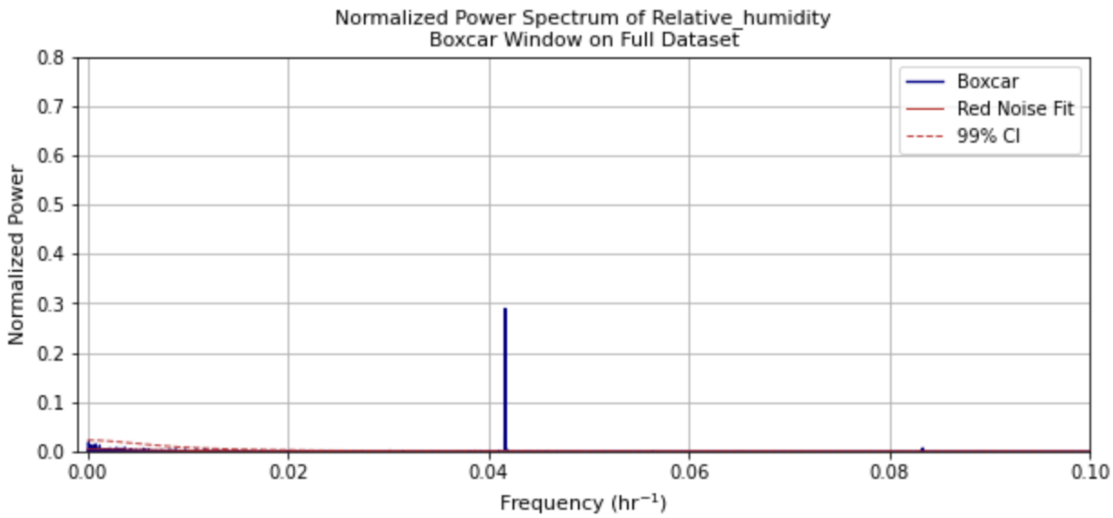
3) Calculate the power spectra using the scipy method. Check that you get the same result as you got using the Numpy method. Next – compare the power spectra obtained using both a Boxcar window and a Hanning window. Assume a window length that is the entire length of the dataset. Do you get the same statistically significant peaks when applying the Hanning window and the Boxcar window? How do they differ? Can you explain why?



The two statistically and physically significant peaks we found in the previous method are also found here. The boxcar method results in higher power spectra in comparison to the Hanning window, which makes sense because the boxcar window is narrower.

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

Relative humidity gives similar 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…

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

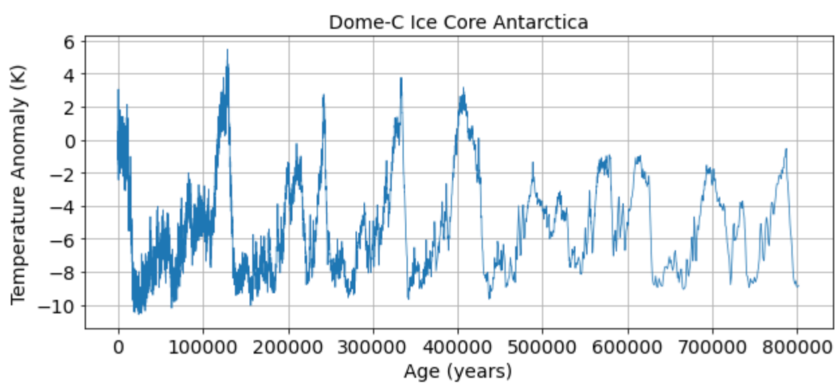
DATA and UNDERLYING SCIENCE:

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

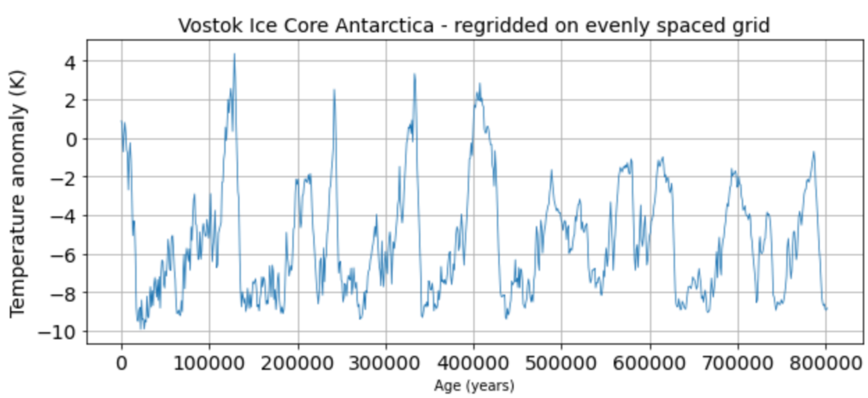
Questions to guide your analysis of Notebook #2:

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.

Before:



After:



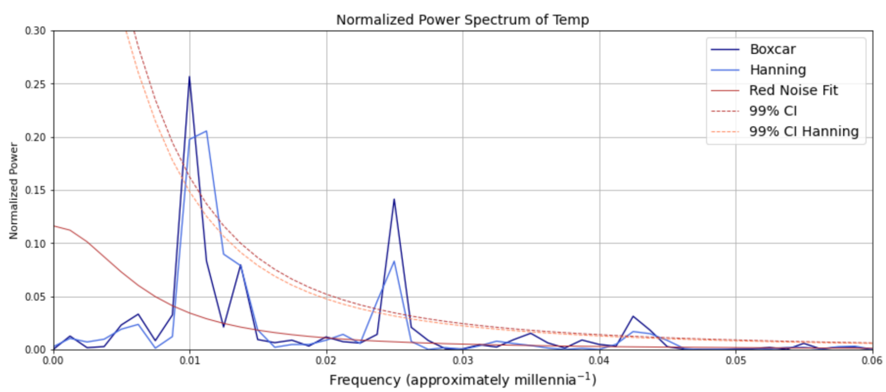
2) Signal and Noise: What is the autocorrelation and e-folding time of your data? What spectral peaks do you expect to find in your analysis and how much power do you think they will have? *Hint: Think back to the Petit 1999 Vostok ice core dataset discussed in class.*

Autocorrelation: 0.96; e-folding time: 25 kry. We expect to see Milankovitch cycles: 100 kyr (eccentricity), 40 kyr (obliquity), 21 kyr (precession).

3) <u>Use Boxcar Window to calculate power spectra:</u> 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 found spectral peaks at 100,328 years, 40,131 years, 23,607 years, and 22932 years, which is what we could expect. I believe that the last two both represent precession, which is known to have a slightly variable frequency. All of these peaks represent the Milankovitch cycles.

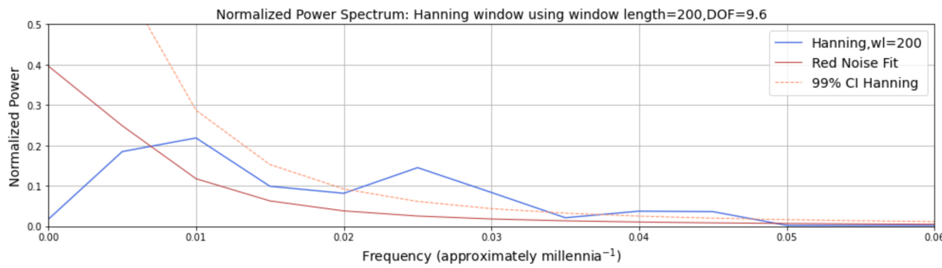
4) <u>Compare Boxcar Window vs. Hanning Window</u>: Calculate the power spectra using the SciPy method. Compare the results obtained using a Boxcar window that is the length of your entire dataset to those obtained using a Hanning window that is the length of your entire dataset. Graph the power spectrum,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

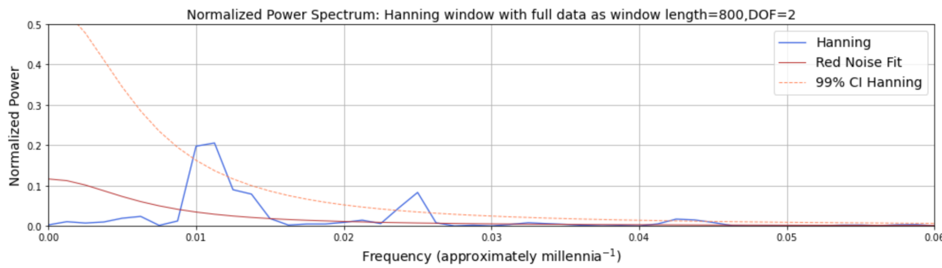


Slightly different (though broadly similar) frequencies are found using a Hanning window: 100,328 years, 89,181 years, 42,244 years, 40,131 years, and 23,607 years. There is a little more spread in the Hanning method, which could likely be due to the window size. These once again represent the Milankovitch cycles. The broadening is the same as what we saw in the Fort Collins temperatures.

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

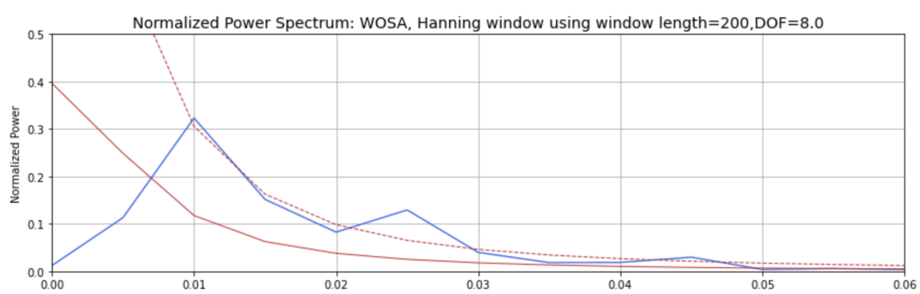
peaks and their statistical significance? Did you find the classic tradeoff between 1) high spectral/temporal resolution but low quality statistics, and 2) high quality statistics but low spectral/temporal resolution?





Changing the window significantly changed how broad the peaks are. The window length of 200 created very broad spectra and many frequencies appear to no longer be significant. For the shorter window, the power is much greater but the peaks are "smeared." In contrast, the longer window has lower power but each peak is more distinct. This is the classic tradeoff mentioned above.

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?



This method results in the greatest power, but the peaks only very slightly exceed the 99% confidence interval. The degrees of freedom are lower for this method (DOF=8) compared to the previous method (DOF=10).