**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?

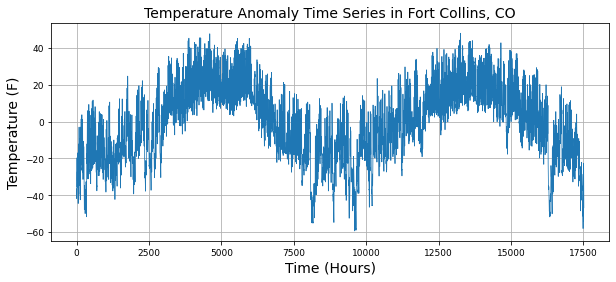
**DATA and UNDERLYING SCIENCE:**

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?

Here are the data for looking:

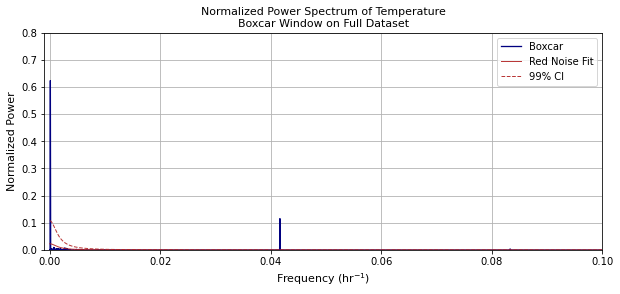


I expect to see spectral peaks associated with seasonal temperature change (frequency 1 yr) and the diurnal cycle (frequency 1 day). I’m not sure what else I’ll see. It looks like the amplitude of the seasonal cycle is 40 degrees F, and the amplitude of the daily cycle is O(40) degrees as well.

The lag-1 autocorrelation is 0.99. Very red. The e-folding time is 101 hours.

1. 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.

The power spectra, red noise fit, and 99% CI are provided in the following plot:

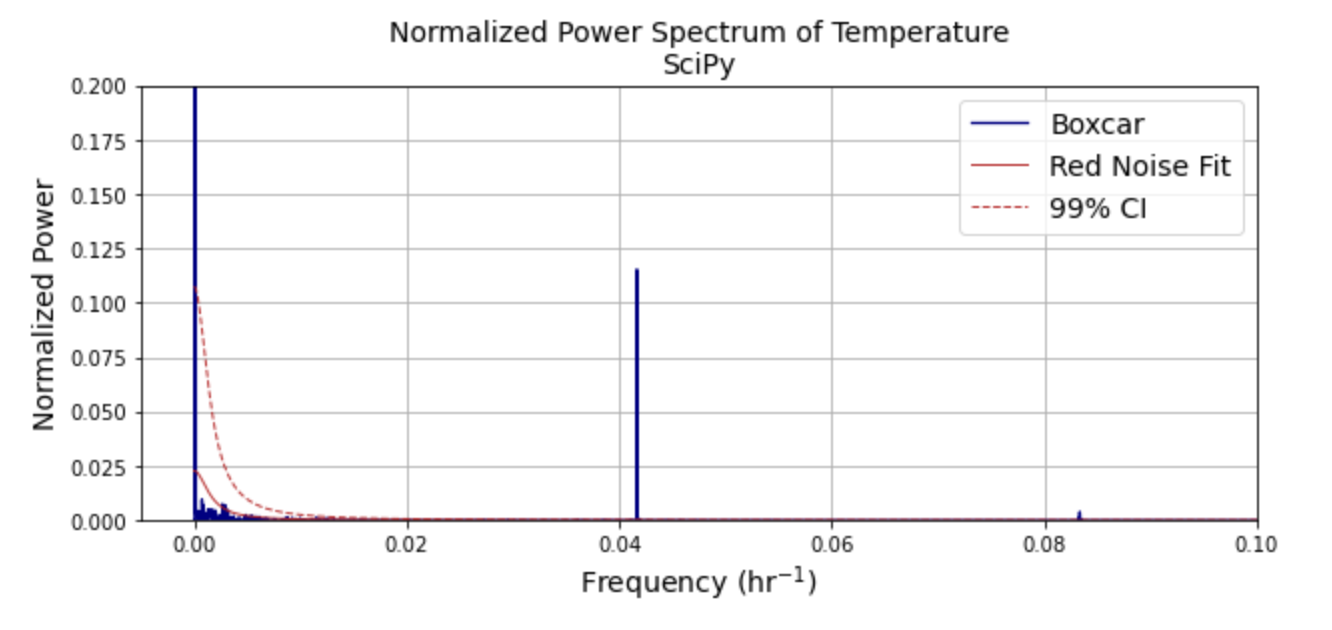


I found the expected daily and annual peaks, as well as some higher-frequency peaks: 0.997 days (probably some sort of artifact, since the physical phenomenon explaining this is still almost certainly the day), and 0.5 days (a harmonic, most likely).

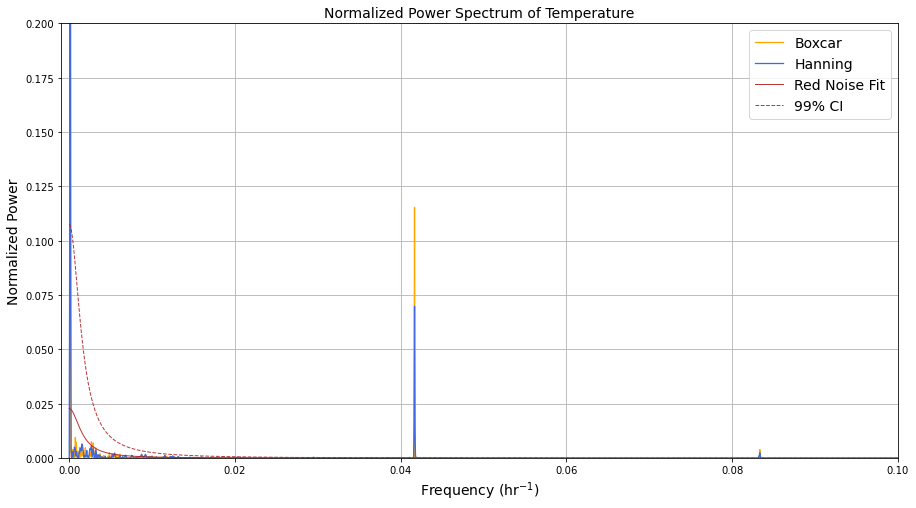
The null hypothesis that I reject for each of these peaks is that the variation is explained by randomness with memory (red noise).

1. 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?

Yup, the spectrum using the scipy method is the same:



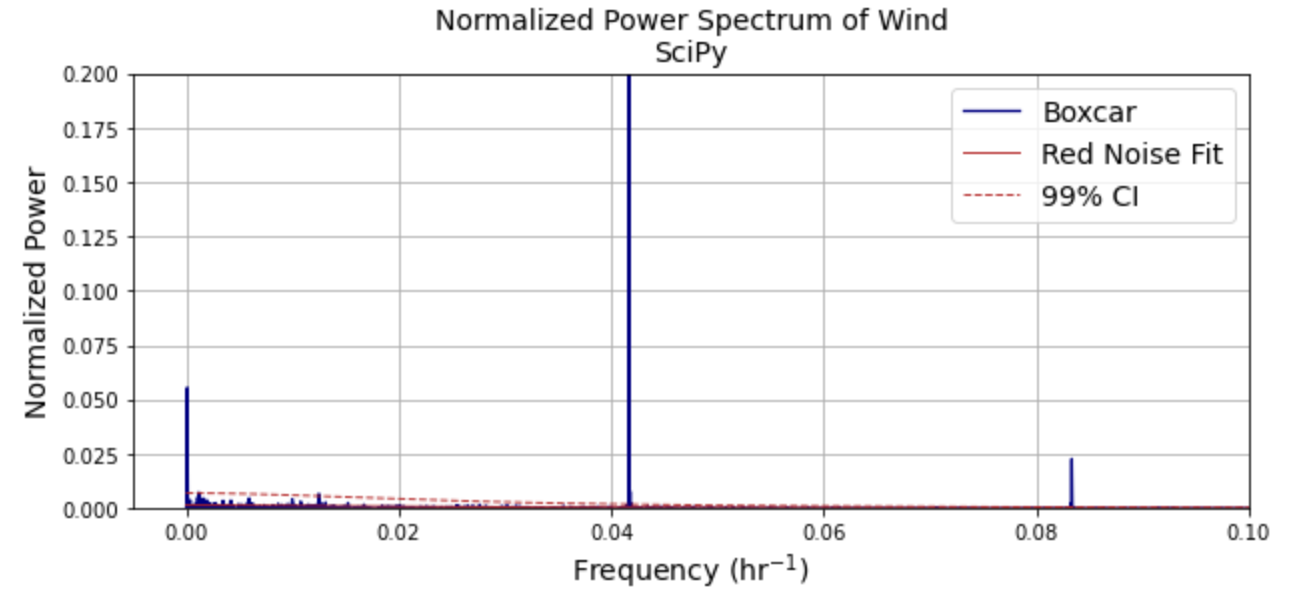
Hanning vs. boxcar:

****

It appears I get a statistical peak at just over 0.04 /hr regardless of the method used. But the boxcar window gives a higher, narrower peak; Hanning gives a more spread one. It appears having more gradual window edge slopes causes some sort of smearing in the frequency domain.

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

When I look at the wind variable, I see a spectral peak at frequency 1/365 /day, 1/day, and 2/day, and 3/day. There are other peaks with no clear physical interpretation. But for the four peaks described, it seems we’re seeing the annual cycle and then diurnal patterns and harmonics thereof.



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…

Yes I am. And I don’t know what it is.

**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:**

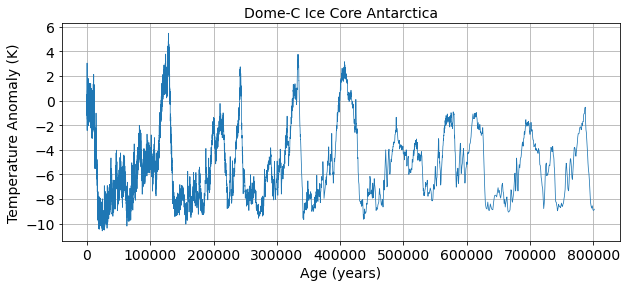
In this notebook, you will perform a power spectral analysis of the temperature record from the Dome-C Ice Core, taken at 75 South and 123 East (Jouzel et al. 2007). The temperature data go back ~800,000 years before present. They are unevenly spaced in time. The data are available on-line here, courtesy of the NOAA Paleoclimatology Program and World Data Center for Paleoclimatology:

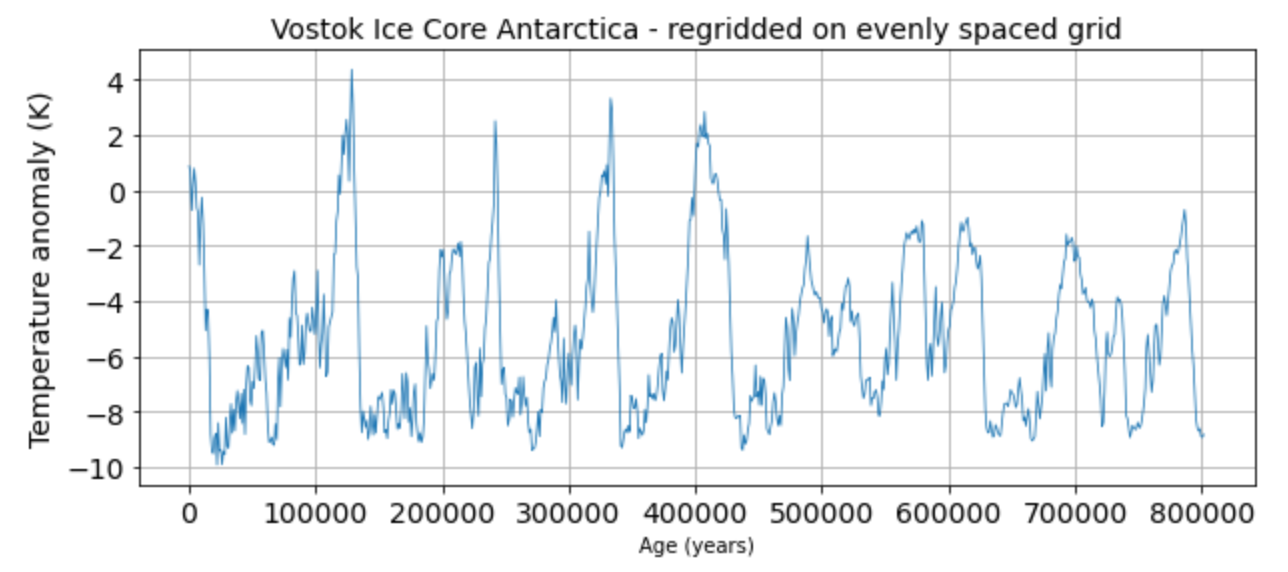
ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/antarctica/epica\_domec/edc3deuttemp2007.txt More information on the data is available at:

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.

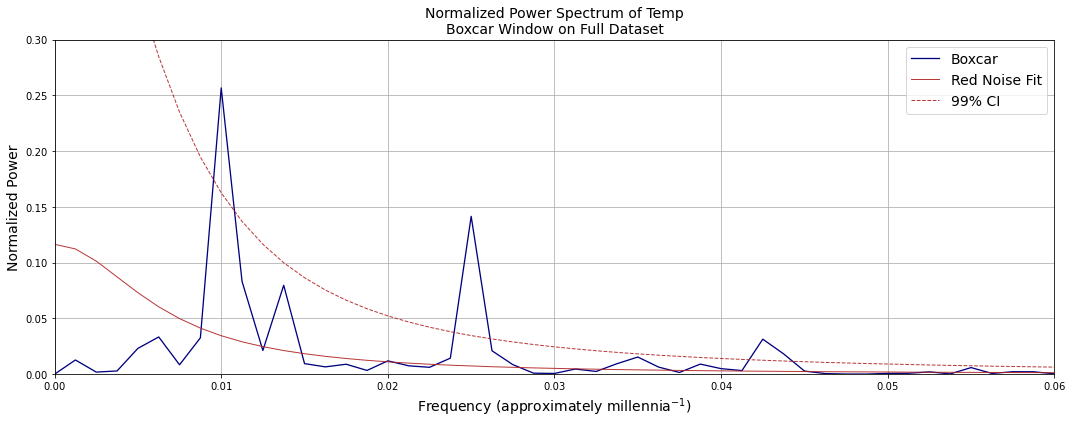
Looks like it worked; these look the same except for obvious Nyquist exclusions:  




1. 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.*

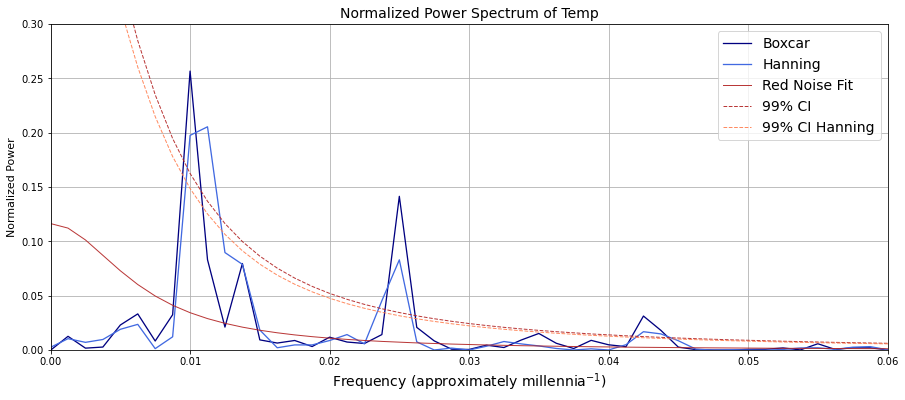
The lag-1 autocorrelation is 0.96. The e-folding time is 25 timesteps. It looks like there’s a major component around period 75,000 years, and another (possibly) around 25,000 years. These guesses are without looking at Petit et al. I still haven’t learned how to guess magnitude.

1. 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?



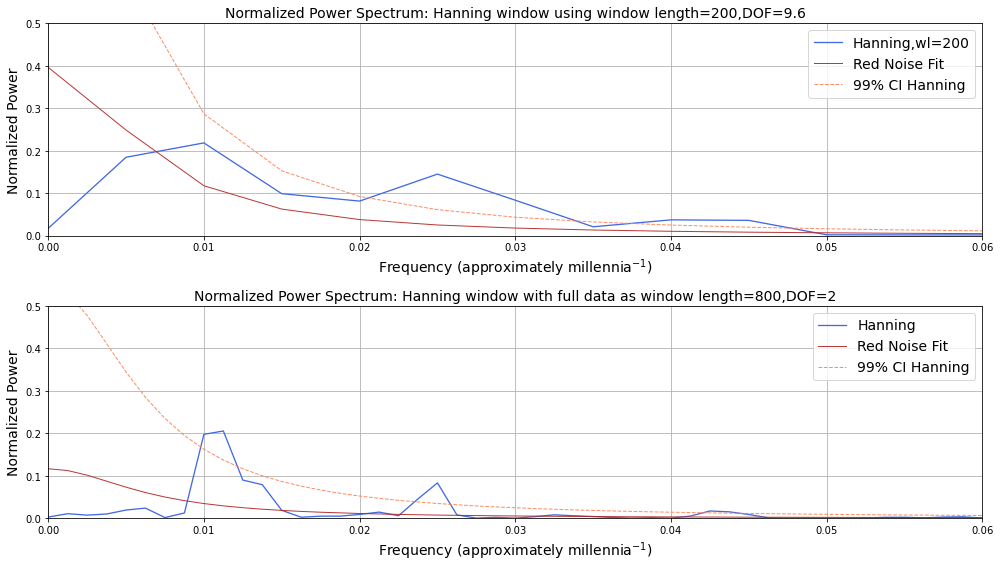
It looks like 0.01 1/millennia, 0.024 1/millennia, and 0.045 1/millennia are statistically significant peaks at the 99% confidence level. These are the frequencies of anomalies in Earth’s orbit.

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?



The Hanning window finds the same statistically significant peaks, but it suggests less certainty as to the frequencies of the peaks. This is consistent with what I found in the first notebook. I think the peaks correspond to the same orbit anomalies.

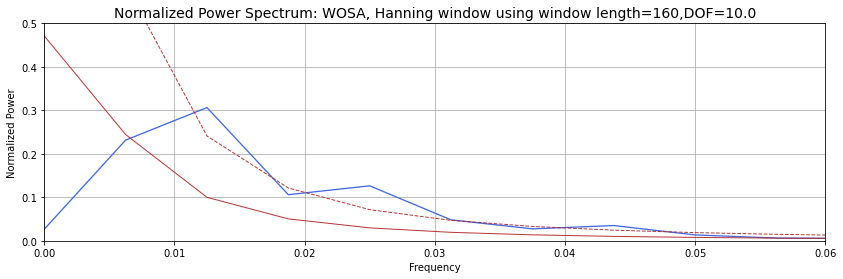
1. Hanning Window with different window lengths: Using the SciPy method, compare the power spectra obtained using Hanning window with different window lengths. Graph the power spectra, the red noise fit to the data, and the 99% confidence interval. Did you find any statistically significant spectral 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?



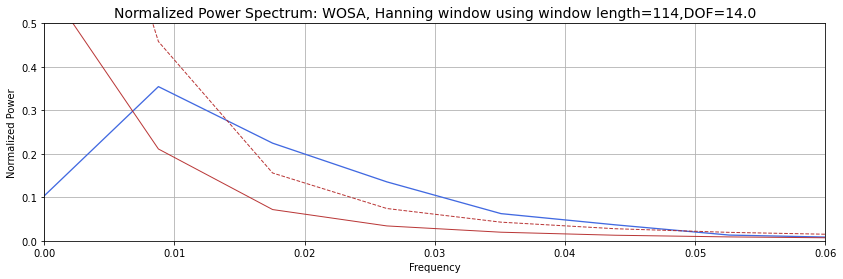
With window length = 200, the only statistically significant peaks are the two higher-frequency ones. The window is too short to give statistical significance for the 100,000-year period oscillation. With window length = 800, the lowest-frequency peak is statistically significant. The frequency of the peaks is also better defined. I’m not seeing the “classic” phenomenon of low quality statistics here.

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?

I tried using window lengths of 1/3, ¼, and 1/5 the data length. Thanks for that formatting, Word. Each detected the three statistically significant peaks I’ve been looking at, but the appearance of the power spectrum graph is less convincing.



When I divide the data into 7 chunks, all becomes chaos:



I don’t have the foggiest notion how to interpret this. Is this telling me there are statistically significant oscillations…somewhere? But I don’t get to know where?