Gina Jozef
3/15/22

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?

Lag 1 autocorrelation: 0.99
e-folding time: 100.92 hours

I would expect to find a spectral peak at 24 hours (1 day), since we would expect to see the diurnal cycle. I would also expect to find a spectral peak at 8760 (365 days), since we should also see the seasonal cycle with two years of data. We hypothesize that the peak representing the seasonal cycle will have a higher power than that representing the diurnal cycle because the change in surface temperature from summer to winter is typically greater than that from one day to night.

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

Gina Jozef
3/15/22

Barnes and Hartman notes to make sure all of the equations and functions in the notebook are working as you expect them too.

Figure 1 below shows the power spectrum with Boxcar window and the red noise fit with 99% confidence interval (CI). Figure 2 shows three panels zoomed in on the peaks so we can see them more clearly. The statistical significance is assessed by comparing the observed power spectrum to that of a red noise time series (the null hypothesis is that the observations are red noise). Given this, technically, four statistically significant spectral peaks are found (the power of the peak is greater than the 99% confidence interval on the red noise fit). The first statistically significant spectral peak has a frequency of 0.0001 and a power in exceedance of the 99% CI of 0.52. This in days is 365, so this represents the annual cycle of temperatures. The next spectral peak has a frequency of 0.04167 and a power exceedance of 0.115. This in days is 1, so this represents the daily cycle of temperature. The third peak has a frequency of 0.04178 and power exceedance of 0.001. This in days is 0.997. We theorize that this is not a real significant peak, but rather a side lobe of the daily cycle which shows up as a result of the boxcar window. Lastly, there is a peak with a frequency of 0.083 and exceedance of 0.00383. In days, this is half a day. I wouldn't expect to see a significant peak with a frequency of half a day, so this is curious.

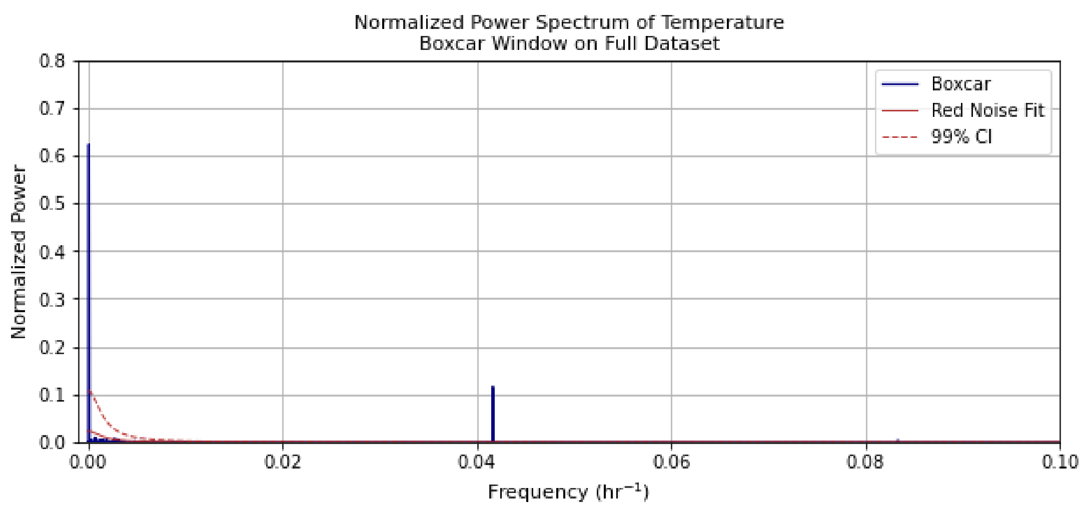

Figure 1: Numpy method - Power versus frequency fit with a boxcar window (blue), along with the red noise fit (red) and the 99% confidence interval (CI) on the red noise fit (red dashed).

Gina Jozef
3/15/22

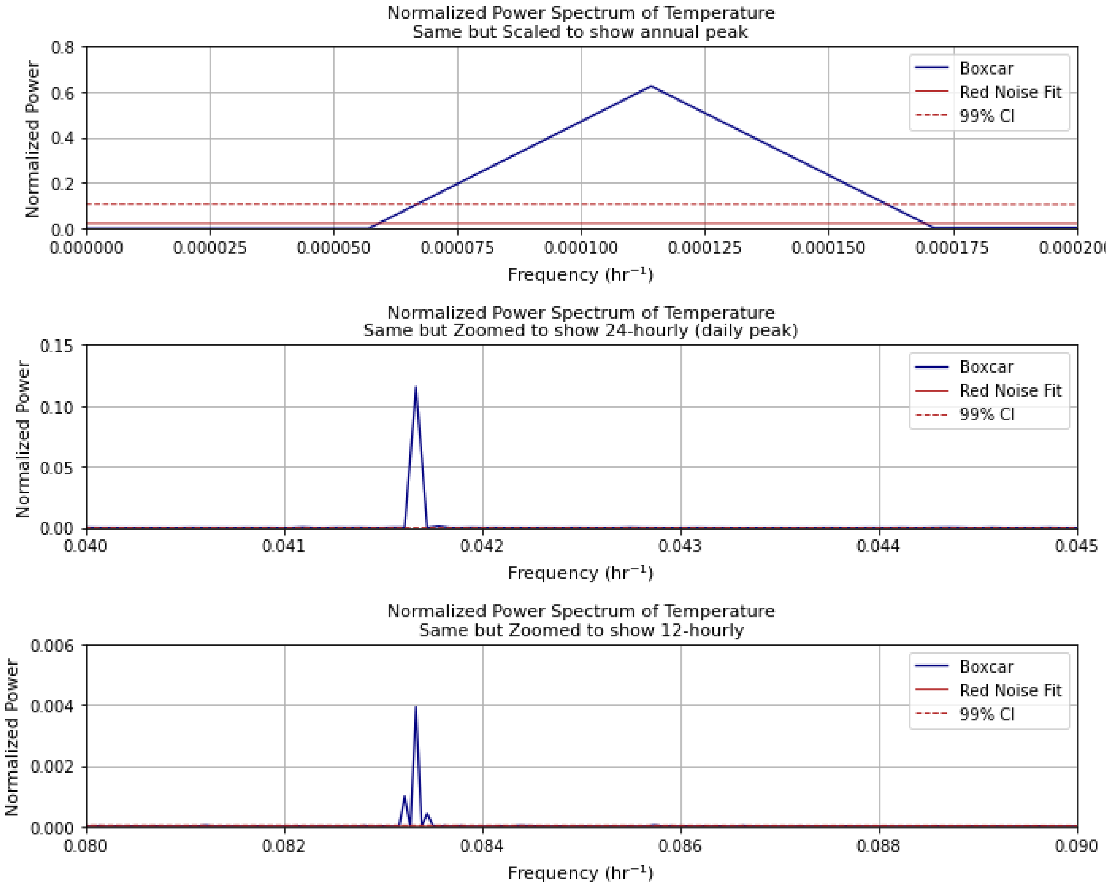

Figure 2: Numpy method - Power versus frequency fit with a boxcar window (blue), along with the red noise fit (red) and the 99% confidence interval (CI) on the red noise fit (red dashed) zoomed in on the annual cycle peak (top), the daily cycle peak (middle) and the half day peak (bottom).

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?

Calculating the power spectrum with the scipy method gives the same result (shown in Figure 3 below).

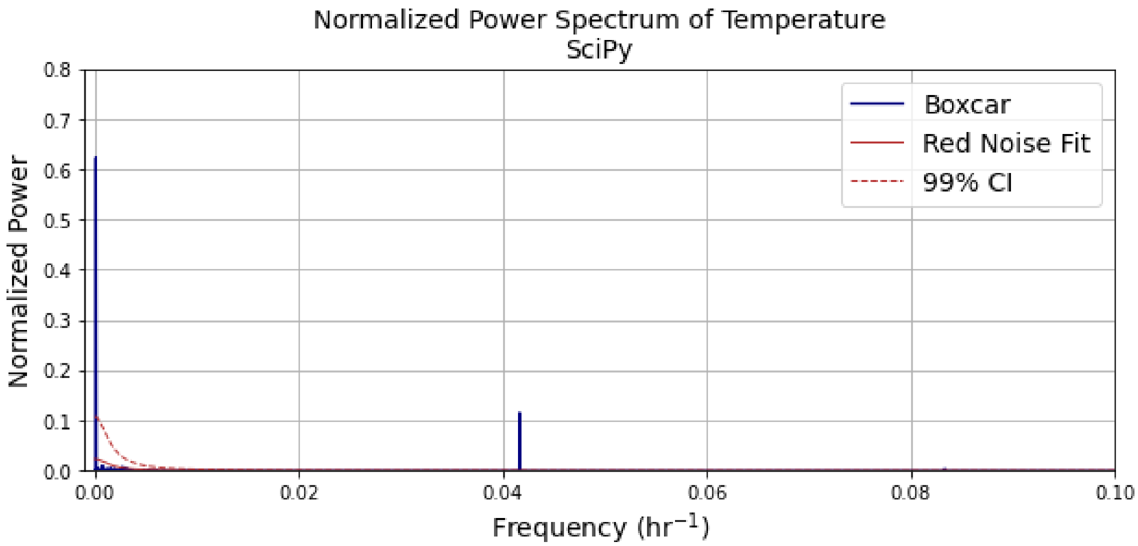

Figure 3: Scipy method - Power versus frequency fit with a boxcar window (blue), along with the red noise fit (red) and the 99% confidence interval (CI) on the red noise fit (red dashed).

Comparing the power spectra obtained using both a Boxcar window with that obtained using a Hanning window, we get the below results in Figure 4 (shown zoomed in in Figure 5). We get three out of four of the same statistically significant peaks when using the Hanning window versus the boxcar window. The missing peak in the Hanning method is that with a frequency of 0.997 days. This further supports that this peak was simply a side lobe resulting from the boxcar window, and the fact that we don't see it with the Hanning method gets rid of that problem.

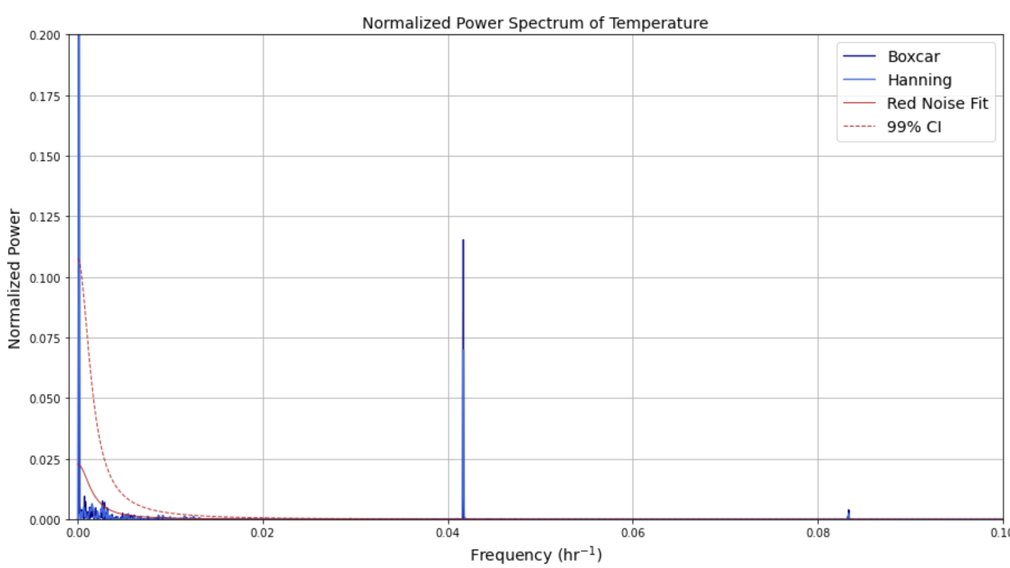

Figure 4: Scipy method - Power versus frequency fit with a boxcar window (dark blue) and Hanning window (light blue) along with the red noise fit (red) and the 99% confidence interval (CI) on the red noise fit from each window method (red dashed).

Gina Jozef
3/15/22

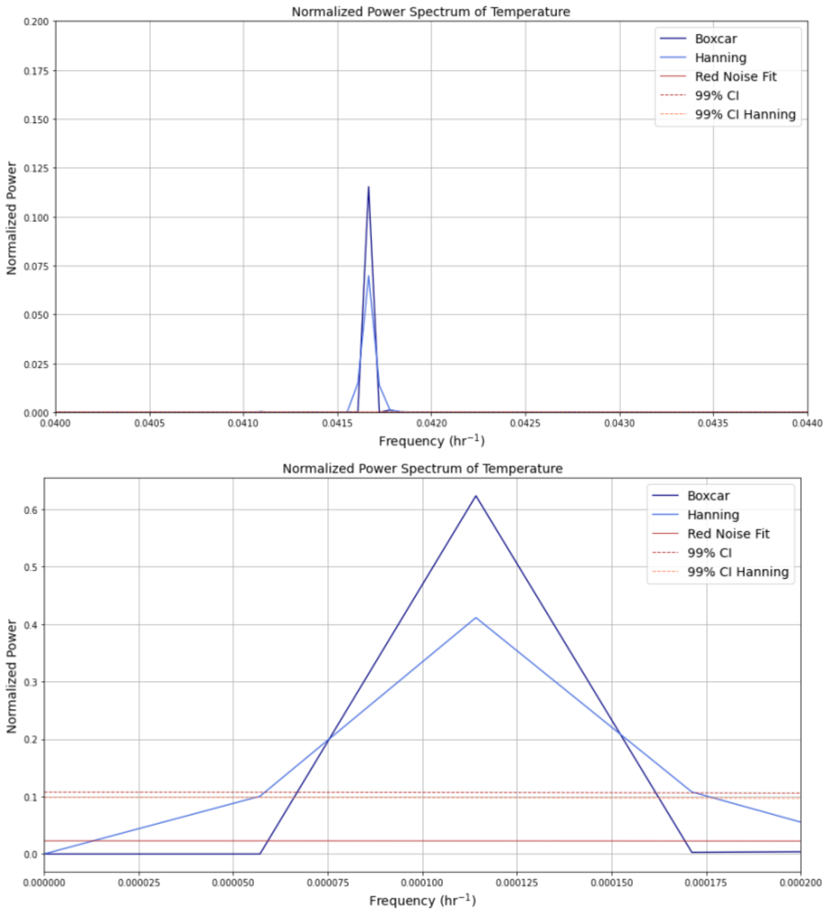

Figure 5: Scipy method - Power versus frequency fit with a boxcar window (dark blue) and Hanning window (light blue) along with the red noise fit (red) and the 99% confidence interval (CI) on the red noise fit from each window method (red dashed) zoomed in on the annual cycle peak (top) and the daily cycle peak (bottom).

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

Using wind direction, we get statistically significant peaks at 365 days, 182.5 days, 11.6 days, 6.7 days, 1 day, and 0.5 days. The peak with the most power is that with a frequency of 1 day. This is likely reflecting the diurnal cycle of wind direction that occurs with mountains. (notebook is ATOC5860_applicationlab4_fft_christman-winddirr.ipynb).

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…

We are seeing significant power at 12-hour frequencies. Very interesting…

Gina Jozef
3/15/22

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.

Given Figure 1 (original data) and Figure 2 (interpolated data), the re-gridding worked as expected.

Gina Jozef
3/15/22

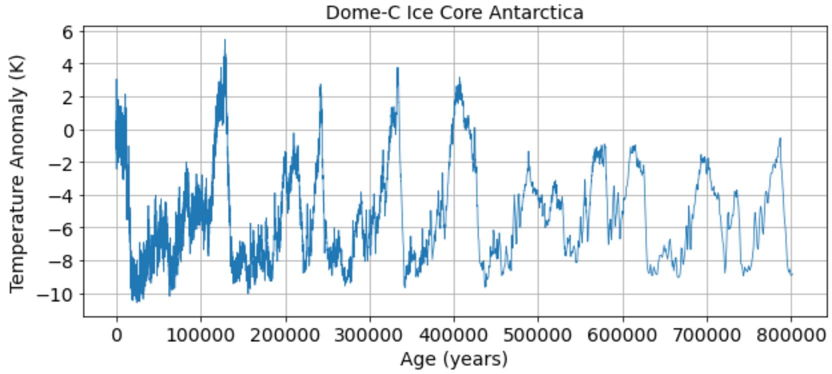

Figure 1: Original time series of temperature anomaly.

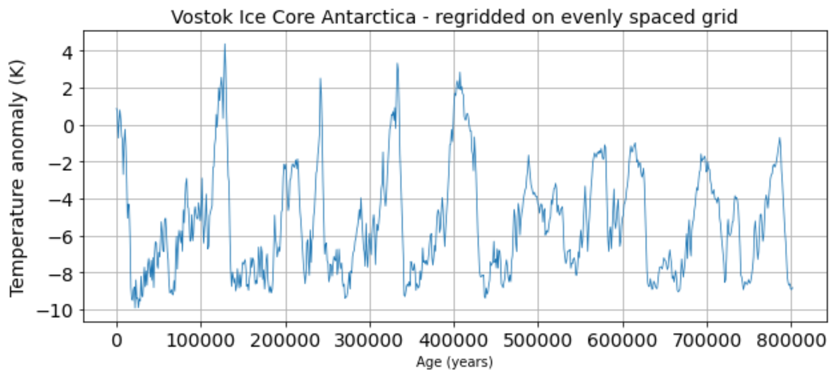

Figure 2: Re-gridded time series of temperature anomaly.

2) <u>Signal and Noise:</u> 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.*

Lag 1 autocorrelation: 0.96
e-folding time: 25,000 years

The spectral peaks we expect to see are those caused by the Milankovitch cycles. These would be peaks with a period of 19-24,000 years which represents change in precession (which direction the earth is tilting), a period of 41,000 years which represents change in obliquity or tilt of the earth (the angle the earth is tilting), and a period of 100,000-413,000 years which represents change in eccentricity (the shape of earth's orbit around the sun). We expect the peak associated with obliquity to have the most power, as this cycle is known to have the biggest effect on climate by varying summer insolation in northern high latitudes.

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?

Figure 3 below shows the power spectrum of temperature anomaly from the ice core dataset. There are four statistically significant peaks. They are:

1. Frequency of 0.01, signifying 100,328 years
2. Frequency of 0.025, signifying 40,131 years
3. Frequency of 0.0425, signifying 23,607 years
4. Frequency of 0.044, signifying 22,932 years

The first peak represents eccentricity, the second peak represents obliquity/tilt, and the third and fourth peaks represent precession.

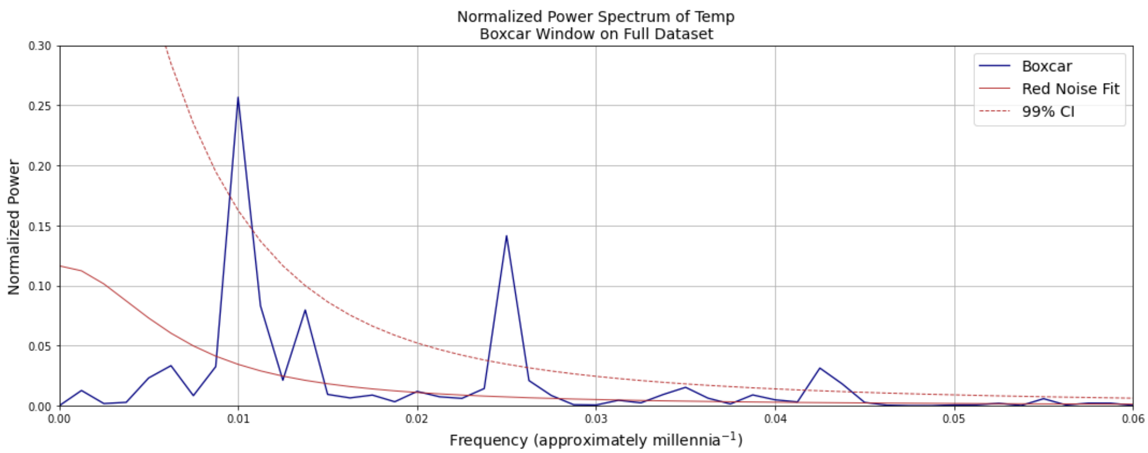

Figure 3: Numpy method - Power versus frequency fit with a boxcar window (blue), along with the red noise fit (red) and the 99% confidence interval (CI) on the red noise fit (red dashed).

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?

Figure 4 below shows the power spectrum using the boxcar method as well as the Hanning window. The Hanning method finds the same statistically significant peaks, however the power of each peak is less. The peak representing precession is only slightly above the 99% confidence level, when using the Hanning window. The peaks from the Hanning window are also broader than the peaks from the boxcar method. Because of this, in the case of the first peak, where it appears that there is a second (though not significant) directly following the first peak ,these are combined into one connected peak when using the Hanning window. Since the timings of the Milanovich cycles span a range, having these broader peaks may better represent the phenomenon. The intuition gained from the Dome-C records is different than that from the Fort Collins temperatures, as the Fort Collins data was telling us about daily and annual climate cycles, whereas the Dome-C records tell us about climate cycles that occur on the scale of thousands to hundreds of thousands of years.

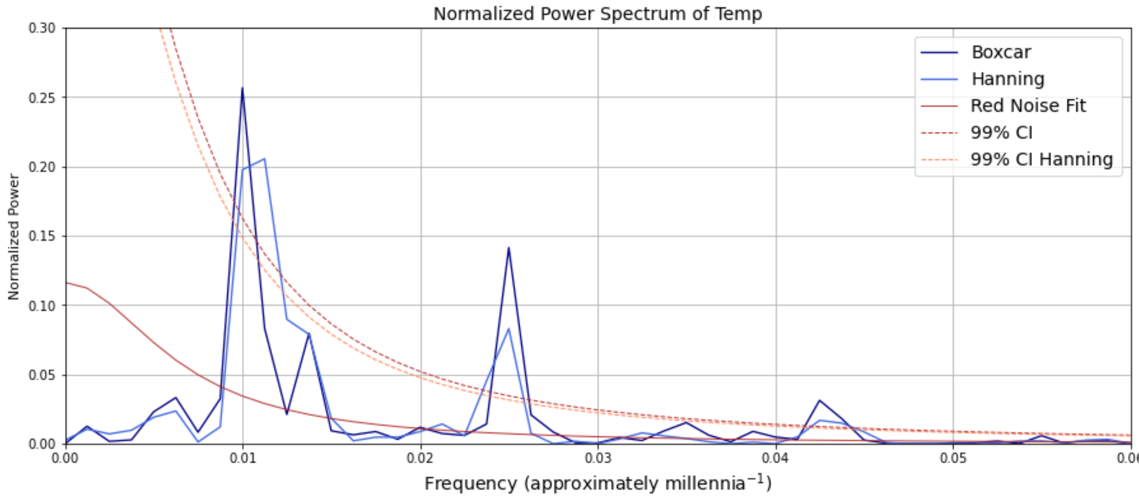

Figure 4: Scipy method - Power versus frequency fit with a boxcar window (dark blue) and Hanning window (light blue) along with the red noise fit (red) and the 99% confidence interval (CI) on the red noise fit from each window method (red dashed).

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?

Figure 5 shows the power spectrum obtained using the Hanning window with different window lengths, where a shorter length of 200,000 years is compared with a window length of 800,000 years (the entire time series). Using the 800,000 year window, we obtain the same results as are shown in Figure 4, since the window is the length of the entire time series. When using the shorter window length, we lose the signal from eccentricity (the first peak). This suggests that using too small of a window length can wash out the signal from lower frequency phenomenon. In this case, the tradeoff between spectral/temporal resolution and the quality of the statistics is in favor of having high spectral/temporal resolution and low-quality statistics.

Gina Jozef
3/15/22

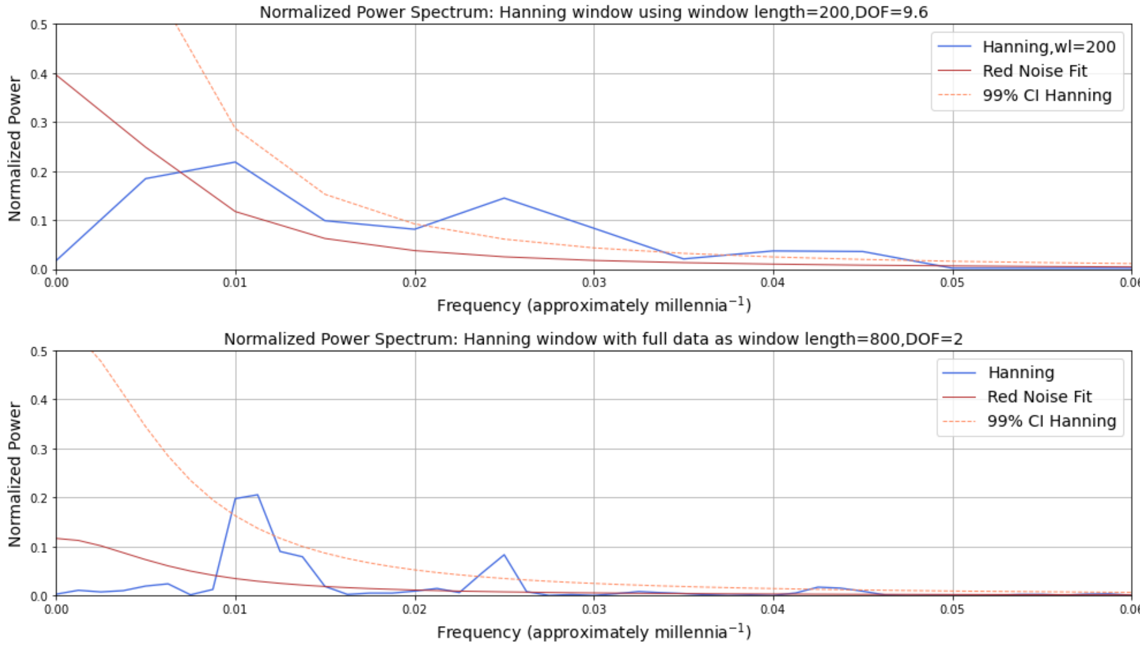

Figure 5: Scipy method - Power versus frequency fit with a and Hanning window of length 200,000 years (top) and 800,000 years (bottom).

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?

Figure 6 below shows the results of applying Welch's Overlapping Segment analysis and using a window length of 200,000 years. Using the WOSA method, we can again see all three significant peaks, though the first and third peak do not exceed the 99% confidence interval by much. However, this now is a good balance between identification of the spectral peaks and statistical significance. We also show the results of the WOSA analysis using a shorter window (Figure 7) and longer window (Figure 8) below. These results suggest that using too short of a window can wash out the difference between spectral peaks, and using a longer window can increase the significance of peaks, while sacrificing the quality of the statistics.

Gina Jozef
3/15/22

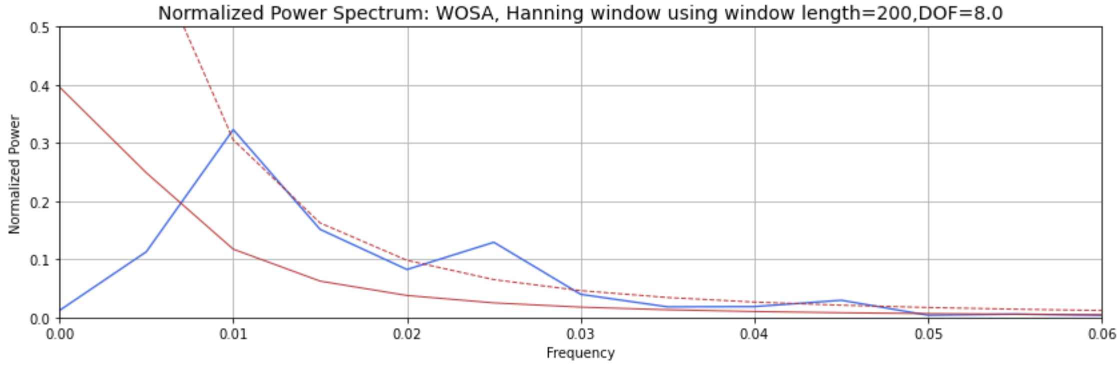

Figure 6: Scipy method - Power versus frequency fit with a and Hanning window of length 200,000 years, applying Welch's Overlapping Segment analysis.

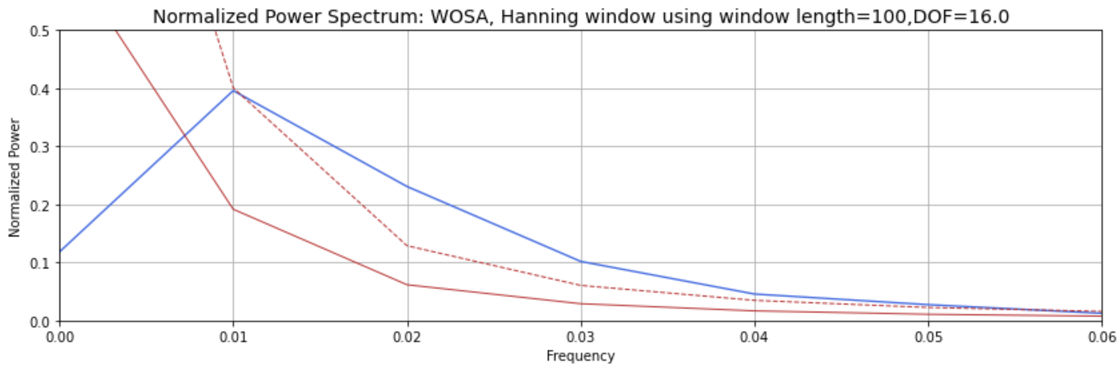

Figure 7: Scipy method - Power versus frequency fit with a and Hanning window of length 100,000 years, applying Welch's Overlapping Segment analysis.

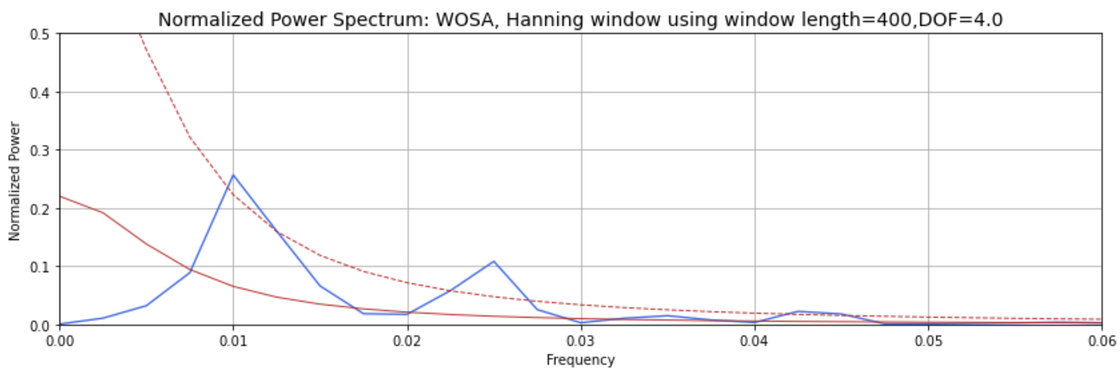

Figure 8: Scipy method - Power versus frequency fit with a and Hanning window of length 400,000 years, applying Welch's Overlapping Segment analysis.