Michelle Maclennan

19 October 2020

**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?

The lag-1 autocorrelation is 0.99, meaning the data have relatively high memory from 1 hour to the next, which we would also expect from a time series of hourly air temperatures. The e-folding time is 100.92 hours. There’s a clear seasonality to the air temperatures so I would expect a large peak corresponding to seasonal variability, and then also peaks for diurnal temperature cycles and perhaps peaks for weather systems lasting 3-5 days on average.

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 to.

Here is the plot of the power spectrum, the red noise fit and the 99% confidence interval:

Table

Description automatically generated

Here is the same plot using a log scale on the y-axis:

Chart

Description automatically generated

I found three statistically significant peaks: one every with a period of 365 days (annual), the second with a period of 1 day (there’s another with a period of 0.99 days) and the third with a period of 0.5 days. The first represents annual frequency, the second daily frequency, and the third diurnal (12-hourly) frequency.

A picture containing table

Description automatically generated

For the statistical significance testing, we are trying to determine where there are statistically significant peaks in the power spectrum. The null hypothesis is that everything is red noise. The peaks that exceed the 99% confidence interval for red noise can be determined to be statistically 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?

Using scipy, you still get the same peaks as before. The power spectrum using a Hanning window is far smoother than the Boxcar window, but both still have the same peaks in power. This is because the window acts as a way to weight the data in performing the Fourier transform. The Hanning window weights the data on a curve, whereas the Boxcar window uses a step function to weight the data in the Fourier transform. The Hanning window means each peak still contains the same area underneath if you were to integrate it, but the peaks are slightly lower at any given frequency.

**Graphical user interface, table, Excel

Description automatically generated**

Same plot but with a log scale on the y axis:

**Chart, bar chart

Description automatically generated**

Smoothing by the Hanning window seen when zooming into individual peaks:

Chart, line chart

Description automatically generated

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

**Notebook #2 – FFT analysis using Dome-C Ice Core Data**

**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.

Here are the data before regridding – they are not evenly spaced in time (many more measurements closer to Age 0):

Chart

Description automatically generated

Here are the data after regridding – now they are spaced evenly in time:

Chart

Description automatically generated

It seems to have worked!

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 autocorrelation is 0.96 and the e-folding time of the data is 25-time steps or about 25,000 years that the memory will decrease by 1/e. With the re-gridding of the dataset, the time steps occur every 1003 years, so these patterns are primarily geared at identifying the Milankovitch Cycles. I assume the most power will be associated with eccentricity (period of 100,000 years), then obliquity (41,000 years), and finally precession (26,000 years).

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?

Chart, line chart

Description automatically generated

I found four statistically significant peaks. The first peak has a frequency of 0.01 millenia-1 or 100,000 years, corresponding to eccentricity or the shape of Earth’s orbit. This peak has the highest power of about 0.25. The second peak has a frequency of 0.025 and a period of 41,000 years, so this corresponds to obliquity (the angle of Earth’s axis with respect to the orbital plane). The 3rd and 4th statistically significant peaks represent the same Milankovitch cycle of precession, with a period of 26,000 years. This is a particularly wide peak, which is why it is associated with two statistically significant frequencies exceeding the 99% confidence interval.

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?

Chart, line chart

Description automatically generated

The same peaks are found here corresponding to eccentricity, obliquity, and precession in both the Boxcar and Hanning methods. However, as we found for the Fort Collins temperatures, the Hanning method yields lower and wider peaks than the Boxcar method. With the Hanning method, the third peak barely exceeds the 99% confidence interval. Because the peaks are wider using the Hanning method, it provides slightly less accuracy in determining the actual frequency or period associated with peaks in power. The flaw in Boxcar is that it may produce a false peak, for example in the eccentricity case Hanning shows one peak for the 100,000-year cycle but Hanning joins the two as one peak, which we know (given outside information) to be more representative of the true Milankovitch cycles.

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?

Changing the length of the Hanning window decreases the number of spectral estimates but increases degrees of freedom. Therefore, you many get higher statistically significance by decreasing the window, but you may lose the peaks.

As you shorten the length of the window, you decrease the sample and increase the degrees of freedom. Because you decrease the sample, the peaks broaden, but at the same time the statistical significance increases due to the higher degrees of freedom. Thus, window length is a tradeoff between accuracy and statistical significance. For example, when you reduce the window length you lose the first peak entirely. This is because the dataset is only 800,000 years long, so the lowest frequency period only occurs about 8 times during the entire period and decreasing the sample massively decreases confidence in the location of this peak.

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?

WOSA overlaps the chunking windows to create more windows than by using non-overlapping windows alone, thereby giving you two additional degrees of freedom per window. Using this method with Hanning, it increases the degrees of freedom and thus the statistical significance of the peaks, but at the sample time you can still use a smaller window length. In applying this method to the ice core dataset, the peak at 100,000 years returns, even though it was not visible with the window length from the previous question, because you’re no longer sacrificing as much data from a cycle with very low frequency. It’s also easier to identify the peaks combining window length with Hanning and WOSA, thereby increasing the spectral accuracy compared to the previous question.