ATOC5860 - Homework 4 - due March 17, 2022

Please send your homework to Jen/Prof. Kay on Slack as a direct message.

Please Name Your Homework Files: "ATOC5860_HW4_LastName.pdf, .html, .ipynb"

Your submissions should include: 1) A .pdf document with responses to the questions below, 2) Your code in both .ipynb and .html format.

Show all work including the equations used (e.g., by referring to the Barnes Notes).
Write in complete, clear, and concise sentences.
Eliminate spelling/grammar mistakes.
Label all graph axes. Include units.
Report values using appropriate rounding.

Problem I) Perform a power spectral analysis on monthly Nino3.4 sea surface temperature (SST) time series (a proxy for ENSO). (50 points total)

Your data: Monthly SST anomalies in the Nino3.4 region from the CESM1 Large Ensemble (http://www.cesm.ucar.edu/projects/community-projects/LENS/, https://doi.org/10.1175/BAMS-D-13-00255.1). You will analyze pre-industrial control runs that represent perpetual 1850 conditions (i.e., they have constant 1850 climate). You will contrast the fully coupled 1850 control run with the slab ocean 1850 control run. The files containing the data are in netcdf4 format and available on the class google drive:
CESM1_LENS_Coupled_Control.cvdp_data.401-2200.nc**,**
CESM1_LENS_SOM_Control.cvdp_data.277-1001.nc

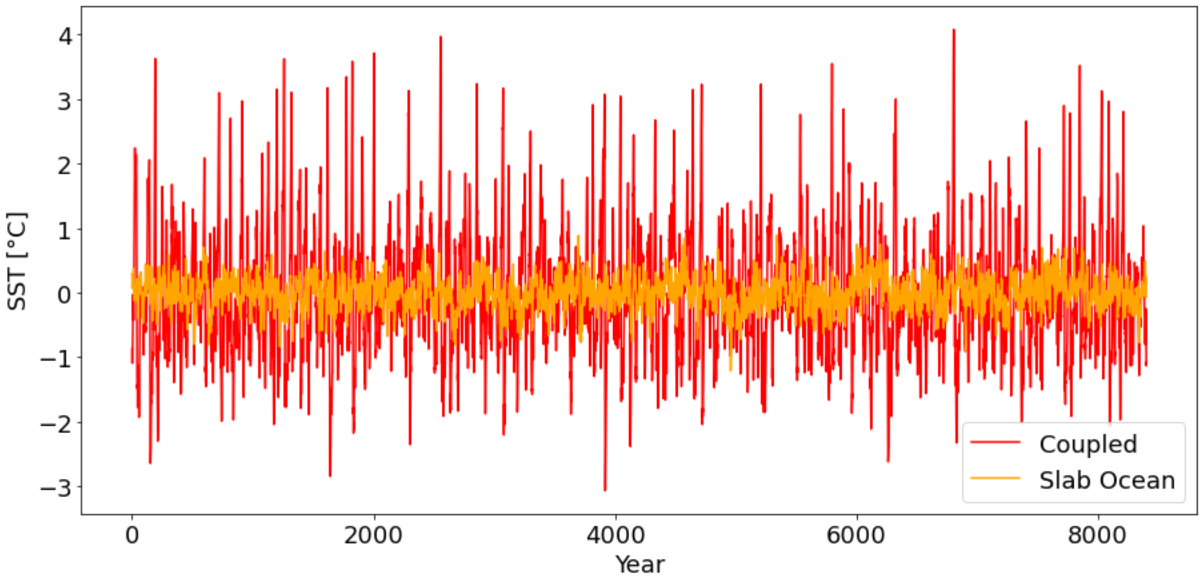
"Munging" (i.e., data pre-processing before doing analysis, yea new words!) for this assignment is minimal. The monthly average SST anomaly in the nino3.4 region is available as a variable called "nino34" in both netcdf files. Munged data from:
http://www2.cesm.ucar.edu/working_groups/CVC/cvdp//data-repository.html

Note: The Climate Variability Diagnostics Package --- the "CVDP" --- (https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014EO490002) might be useful resource to check out for problem 2 or a future homework. List the available variables using your favorite tool to print the file header to the screen (e.g., type "ncdump –h CESM1_LENS_Coupled_Control.cvdp_data.401-2200.nc")

a) Calculate the power spectra of the Nino3.4 SST index (variable called "nino34") in the fully coupled model 1850 control run and slab ocean model 1850 control run. Apply the analysis to the first 700 years of both control runs. Use Welch's method (WOSA!) with a Hanning window and a window length of 50 years. Make a plot of normalized spectral power vs. frequency for both the fully coupled and for the slab ocean. Analyze the statistical significance of spectral peaks using a 99% confidence level and the null hypothesis of red noise. Assume that the factor f_w to compensate for extra smoothing (Eq. 59 Barnes Chapter 4) is 1.0. Your answer should include: 1) a timeseries plot of the original data (Look at your data!!), 2) a plot of the normalized power vs. frequency for the fully coupled model including significance testing vs. a red-noise null hypothesis, 3) 2) a plot of the normalized power vs. frequency for the slab ocean including significance testing vs. a red-noise null hypothesis, 4)

A discussion of your results answering the questions: Is there statistically significant power in either time series and at what frequencies? How do both the amplitude and the statistical significance of spectral peaks in the slab ocean and the fully coupled 1850 control runs differ? (30 points)

The plot below shows the monthly timeseries of Nino3.4 for the first 700 years of the datasets. Both timeseries are detrended since the FFT assumes stationarity of data. However, this step is likely not required for this application as a pre-industrial 1850 control run would not incorporate global warming.



The sample mean is subtracted from each timestamp to obtain the temperature anomalies. These temperature anomalies, warmer or cooler sea surface temperatures, indicate the ENSO phase. Next, the autocorrelation of the each timeseries is calculated at a lag of 1 timestep. This step is performed following Barnes Chapter 2 Equation 68:

$$\gamma(\tau) = \overline{x'(t)x'(t+\tau)}$$

where $\gamma(\tau)$ is the autocovariance, τ is the lag of 1 timestep, and prime values indicate a perturbation from the mean. The autocovariance is normalized to the autocovariance at a lag of 0, or the population standard deviation $\overline{x'^2}$, to retrieve the autocorrelation.

Next we calculate the e-folding time for a red noise timeseries with the same autocorrelation following Hartman Chapter 6 Equation 62:

$$T_e = \frac{-1}{\ln\left(\gamma(\tau)\right)}$$

The autocorrelation and e-folding times of the fully coupled model and slab ocean models are 0.97 and 32 years and 0.93 and 14 years, respectively. We create a power spectrum for this red noise for comparison following Hartmann Chapter 6 Equation 64:

$$\Phi(\omega) = \frac{2T_e}{1 + \omega^2 T_e^2}$$

Next, we compute the critical f-score from a 99% confidence level and 200 degrees of freedom. 200 degrees of freedom are used to create high-resolution accuracy on the frequency domain.

The critical f-score can be compared with the power spectra of the data to determine if peaks in power are statistically significant. When computing the f-score, we assume a null hypothesis that the data is red noise and not periodic. This value is computed from Barnes Chapter 4 Equation 34:

$$f = \frac{S_1^2}{S_2^2}$$

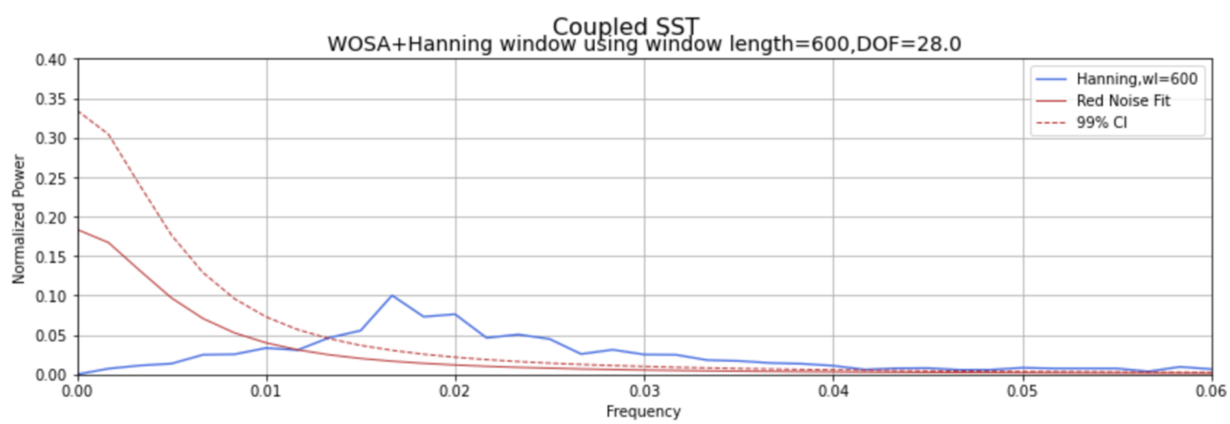
where S_1^2 and S_0^2 are the variances of the original data and the red noise fit, respectively. In order to reject the null hypothesis, the power must exceed the citial value.

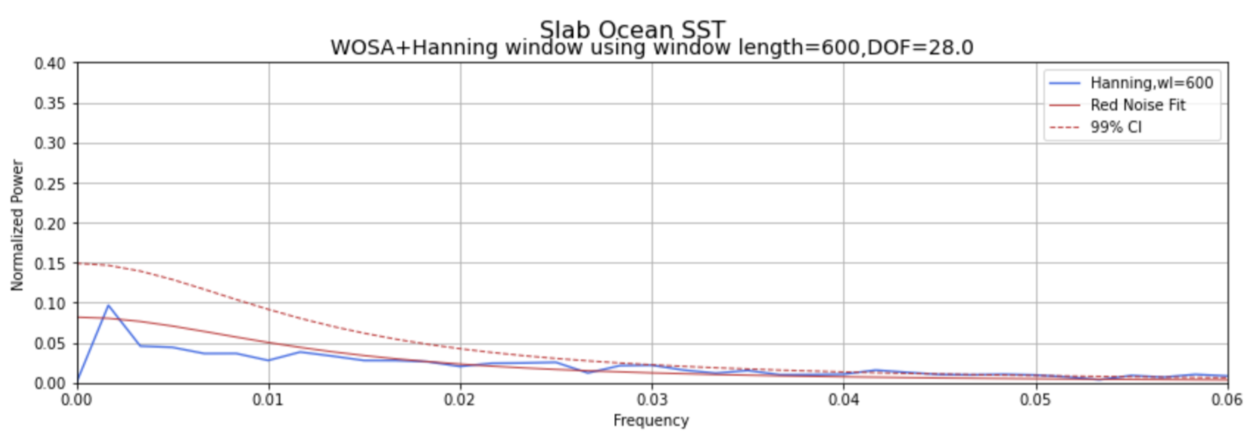
Next we compute the power spectrum for each dataset using the Welch's method and an overlapping Hanning window of length 50 years. We calculate the degrees of freedom from Barnes Chapter 4 Equation 59:

$$DOF = \frac{N}{(50/2)} f_w$$

Where N is the timeseries length of 700, 50 is the window length, and f_w is here assumed to be 1. This leaves a DOF of 28 for both datasets. Finally, we calculate the power using a built in function at each frequency which respresents Barnes Chapter 4 Equation 42:

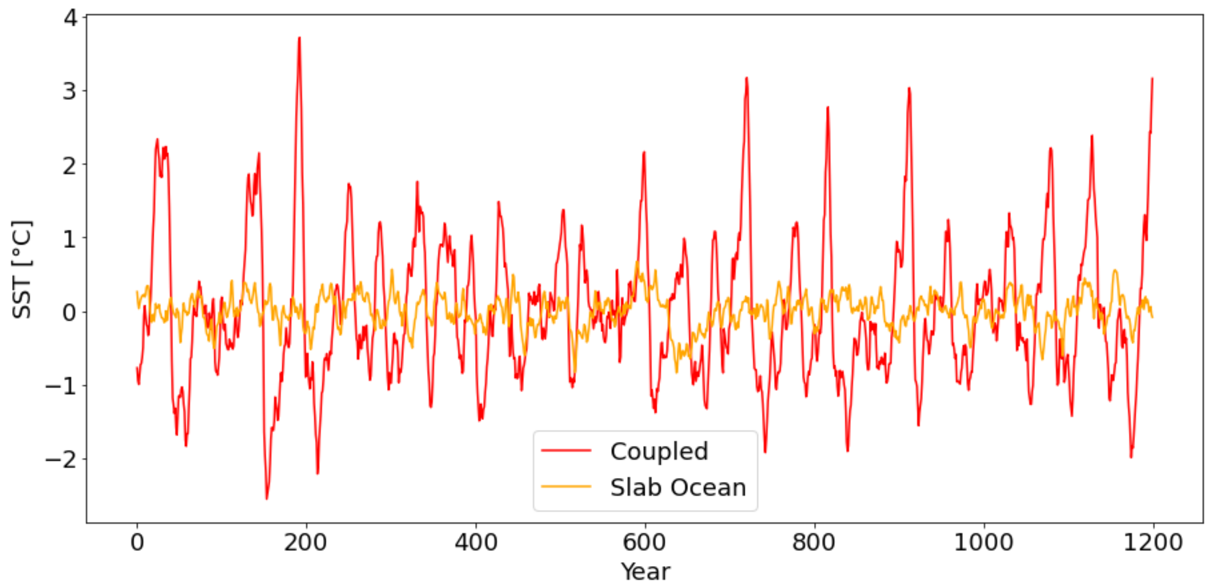
$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} d\omega$$

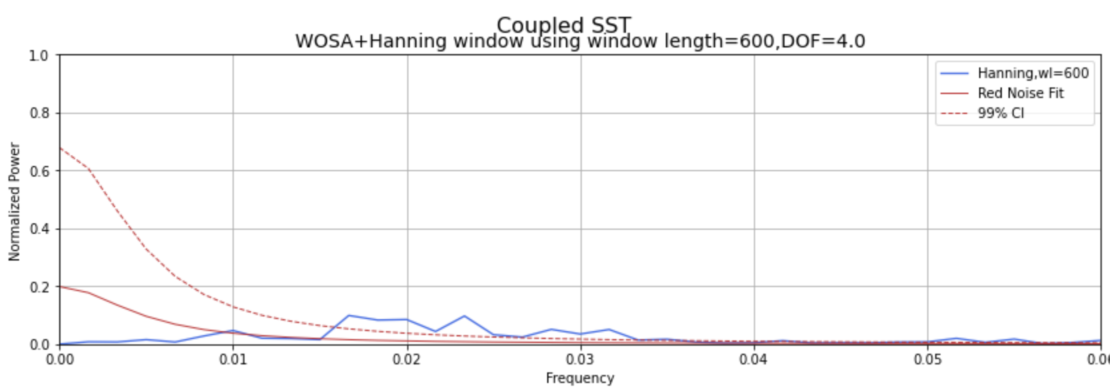


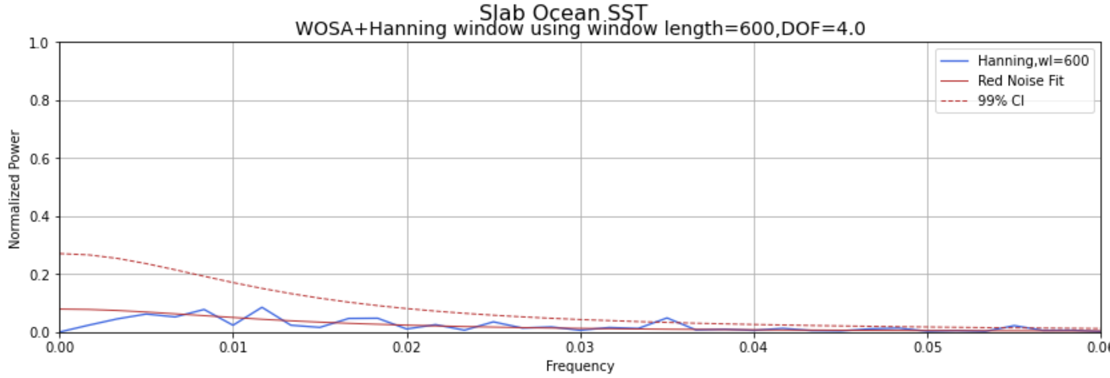


When using the coupled model, we find statistically significant power at frequencies greater than 0.013 month^{-1}. The greatest power, though, is found at the frequency of 0.016 month^{-1}. When using the slab ocean model, the frequencies of statistically significant power range from 0.041 to 0.043 month^{-1}. The greatest power is found at a frequency of 0.0016 month^{-1}.
The amplitude of maximum power in the coupled model is 0.1 while the maximum amplitude in the slab ocean model is 0.096. So, the maximum amplitudes are roughly the same although smaller for the slab ocean model. The confidence interval is higher for the coupled model than for the slab ocean model, reaching a maximum of 0.35 as opposed to 0.18.

b) Redo the analysis in a) for the first 100 years only. Provide the same plots and answer the same questions as a). Then assess - ***How do your results differ?*** **(10 points)**







When using the coupled model, we find statistically significant power at frequencies from 0.016 to 0.033 month^{-1} and those greater than 0.05 month^{-1}. The largest power is found at the frequency of 0.016 month^{-1}. When using the slab ocean model, the one statistically significant frequency is 0.035 month^{-1}. The greatest power is found at a frequency of 0.011 month^{-1}.
When using the slab ocean model, we find statistically significant power at the frequencies 0.035 and 0.055 month^{-1} The amplitude of maximum power in the coupled model is 0.098 while the maximum amplitude in the slab ocean model is 0.085. However, the confidence interval is higher for the coupled model than for the slab ocean model, reaching a maximum of 0.7 and 0.3, respectively.

When using only 100 years of data, the results have changed. In the coupled model, the minimum frequency for statistical significance has increased from 0.013 to 0.016 month^{-1} and the peaks of significant amplitudes has reduced significantly. For the slab ocean model, the opposite is true; the frequency range for statistically significant amplitudes has reduced from roughly 0.041 to 0.035 month^{-1}. The maximum amplitudes of power have reduced somewhat, although the amount is negligible. The maximum 99% confidence interval has increased substantially when using only 100 years of data. For the coupled model, the highest confidence interval increased from 0.35 to 0.7. Similarly, the highest confidence interval for the slab model increased from 0.18 to 0.3.

d) Discuss your results. Provide a physical interpretation for the statistically significant spectral peaks. What do the frequencies that have significant power represent? What do the fully coupled and slab ocean runs have in common and also... How and Why do they differ? ***Hint: A quick refresher on the atmosphere-ocean coupling that contribute to ENSO and skimming Dommenget, D. (2010) doi:10.1029/2010GL044888 might be helpful.*** **(10 points)**
The overall ranges of frequencies between the above problems containing statistically significant power are 0.013 month^{-1}and greater, 0.041-0.043 month^{-1}, 0.016-0.033 month^{-1}, 0.035 month^{-1}, and 0.055 month^{-1}. The lowest of these frequencies corresponds to a timescale of 76 months. The highest frequency corresponds to a timescale of 18 months. Physically, this means that we could expect SST anomalies in the Nino3.4 region to occur at a periodic cycle anywhere from 1.5 to 6.3 years. This finding is close to the actual ENSO cycle, which contains a periodic timescale from about 3 to 7 years. So, our analysis has a slight negative bias.

Both the fully coupled and slab ocean models produce variability in sea surface temperature. The slab model is forced by atmospheric processes including shortwave, longwave, and sensible heat fluxes so ENSO variability is observed. However, the amplitude of ENSO SST anomalies in the slab model are muted since ocean dynamics are not included. For example, atmospheric wind stress could advect cooler sea surface temperatures to a particular region, but the underlying ocean dynamics will not change. So, the inclusion of cold tongue upwelling, as one example, is not represented in the model, making the ENSO cold region less amplified. This could be why the slab model fails to produce large power at frequencies between approximately 0.013 and 0.04 month^{-1} (1 to 6 years) in both a) and b) and why the range of SST anomalies is smaller in the timeseries compared with the coupled model.

Problem II) Apply spectral analysis to a dataset of your choice (50 points)

a) **Provide a thorough description of your data including the reference, the variable (including names and units), the sampling frequency, etc. Describe why you think power spectral analysis may provide useful information about your data. (10 points)**

This analysis will use data output from the Weather Research and Forecasting (WRF) model. The simulation was run at a 2-km horizontal resolution and a 10-m vertical resolution. The data exists at a 10-min frequency encompassing an entire year from September, 2019 to September, 2020. Each variable of interest has dimensions [Time,level,lat,lon]= [52704,17,258,465]. Since this dataset is large enough that it can crash Jupyter, the data is subset to an hourly output frequency, the closest model level to the surface, and to a model grid cell near the coastline.

In the last homework, I computed the first three EOF structures of wind direction. The EOF which explained the most variance was a 70° shift in wind direction near the coast. The average wind direction in this region is parallel to the coast. I presumed in my discussion that this shift, which is nearly perpendicular to the mean flow, was caused by diurnal and seasonal heating contrasts between land and sea. Namely, the temperature gradient caused by differences in surface heating are conducive to sea/land breezes. So, I think this power spectral analysis will provide useful information in determining if the EOF structures I analyzed in the previous homework were indeed caused by diurnal and seasonal variability.

b) **Explain what you expect to see in the power spectra plots. Do this **before** you do the analysis. What is your hypothesis? (There is not any penalty for it being wrong…!) (10 points)**

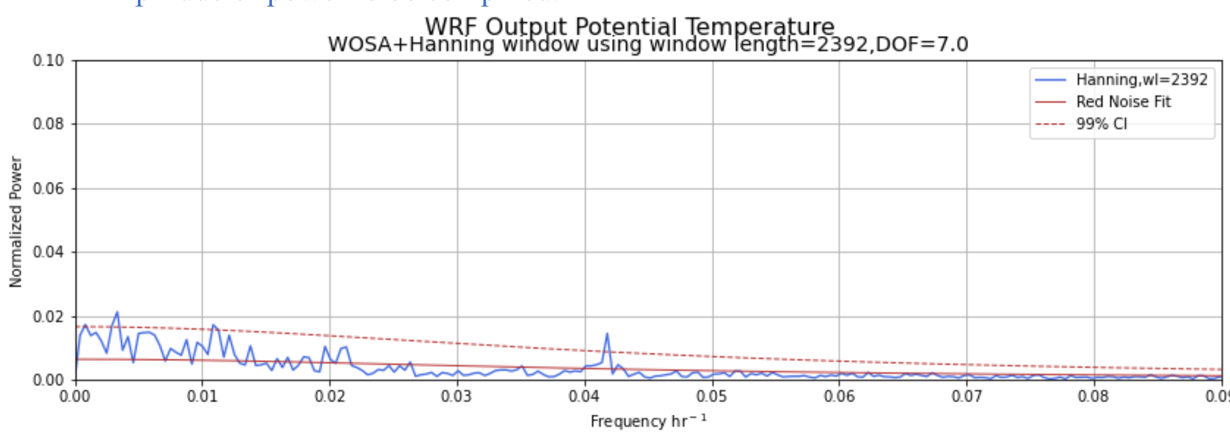
I expect to see power at frequencies corresponding to a 24-hour period. I expect this because land/sea breeze circulations are induced by daytime heating and follow the diurnal cycle. Although the temperature contrast between the land and sea surface also follows a yearly timescale, I only have 1 year of output. So, I don't think this analysis will have power at a frequency corresponding to a yearly timescale.

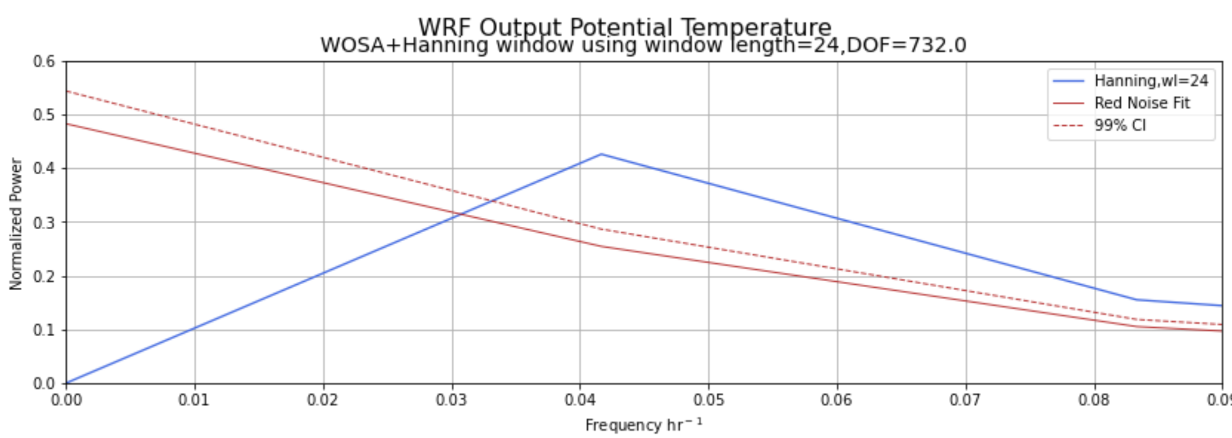
c) **Perform FFT analysis and find the power spectrum. Provide a plot of normalized power vs. frequency. Analyze the statistical significance of any spectral peaks using a 99% confidence level and the null hypothesis of red noise. (20 points)**

The plot below shows the raw wind direction data plotted for the full year from September 2019 to September 2020.

Using the same method as before, the lag-1 autocorrelation is 0.76 and the e-folding time is 4 hours. For this analysis, we use a sliding Hanning window of length 24. We do not care about frequencies higher than the daily scale, as these are likely

For the power spectral analysis, we will test two different window lengths. The first method (top figure) leaves only two windows for use. The second method (bottom figure) uses windows of length 24 hours. We choose the cutoff of 24 since we are not concerned with frequencies higher than a daily cycle leaking to the lower (daily) frequency. This is because wind direction is highly sensitive to thermal gradients and thus typically varies at frequencies controlled by a diurnal cycle or lower frequencies. Furthermore, comparison of a short and long window creates both high temporal resolution and high fidelity in the amplitude of power to be compared.





From the plots above, the first method, using two windows only, results in three distinctive frequencies that have statistically significant power. These frequencies are 0.003 hr^{-1}, 0.01 hr^{-1}, and 0.041 hr^{-1}. The greatest power is found at the frequency of 0.003 hr^{-1} with an amplitude of 0.021. The second plot, which uses 2392 windows of length 24, results in a large range of statistically significant frequencies. This range exists from roughly 0.034 hr^{-1} and greater. The greatest power is found at the frequency of 0.041 hr^{-1} with an amplitude of 0.42.

d) Describe what you found. Were there statistically significant spectral peaks? What do the peaks mean physically? What did you learn? Did you find what you expected to find? (10 points)

Both methods return statistically significant spectral peaks. The first method finds three spectral peaks with high temporal resolution although the fidelity of statistical significance is low. The second method returns only one spectral peak with low resolution but the fidelity of statistical significance is high. Combining the two methods, there is one common frequency of 0.041 hr^{-1}. We can be confident with the finding since this frequency exists both at a high resolution and with a large amplitude and thus high fidelity of statistical significance. This frequency corresponds to a timescale 24.3 hours, or the diurnal cycle. This finding is not surprising since we found in homework 2 that the dominant mode of variability along the coast was a perpendicular veering of the wind. The main cause of this steering is the land/sea-breeze that also follows a diurnal cycle.

Note for grading Problem II. You are analyzing your own data. Since only you know the "right answer", you will be largely graded on how well I can follow your description of the data, the methods, the results, and the conclusions. Keep your code and your explanations simple, clear, and easy to follow. Spend the time to make your code concise, clear, and well documented. Look at your code as an opportunity to re-enforce the understanding you have gained in class as expressed through analyzing your own data.

3) Homework #5 will also require that you analyze a time series dataset of your choice. You will apply more power spectra analysis and filtering to this dataset. A dataset with interesting variations at a wide range of timescales will be the most interesting. I strongly encourage you to use a similar dataset for both Homework #4 and #5. Please describe the dataset you plan to use here. Discuss with me or a classmate if you do not have a dataset in mind. We can help you brainstorm!! (0 points)

For homework number 5, I want to further analyze wind direction. However, instead of looking at wind direction along the coast, I will choose a point in complex terrain. Wind directions in complex terrain can be influenced by the changes in heating along the slope. These changes can depend not only on the position of the sun in the sky (for a diurnal cycle) but which direction the slope faces. This may result in a range of variability on multiple timescales.