Project 4: Spectral Analysis and Signal Processing

Elias Gutierrez[1, a)]
UCLA Physics and Astronomy, Knudsen Hall 475 Portola Plaza, Los Angeles, CA 90095

(Dated: 30 August 2017)

Several calculations were made for the sunspot and sun ray activity data: the calculation of the time period of the dominant nonzero frequency mode in the spectrum for the sun rays and sunspots are 21 years and 67 years, respectively; and the estimated time period of the dominant nonzero frequency mode in the spectrum for cosmic ray cycle and solar cycle are 70 years and 65 years, respectively. Lastly, the correlation coefficient was found at a value of $C_{X,Y}(l=0)2.7\times10^{28}$, and the global minimum shifted to the right where it was not at $l=0$ was approximately at a correlation coefficient of 2.8×10^{28} at a lag time of 7 years.

I. INTRODUCTION

Any periodic function $X(t)$ with period T may be written as a sum of sinusoidal waves:[1]

$$\hat{X}_k = \frac{1}{T}\int_0^T X(t)e^{\frac{-2\pi ikt}{T}}dt, \tag{1}$$

$$X(t) = \sum_{k=-\inf}^{\inf} \hat{X}_k e^{\frac{2\pi ikt}{T}}. \tag{2}$$

The coefficients $\hat{X}_k$ are complex numbers and they define the amplitude and phase of a wave with frequency (find a way to make the f like the math version) $f(k) = k/T$.[1] Equations (1) and (2) give the forward and backward Fourier transforms on the interval $[0, T)$, respectively.[1] When the signal $X(t)$ is sampled using N regularly spaced points ($X_n = X(n\Delta t), \Delta t = T/N$) we have to use the discrete version of the Fourier transform.[1] The most common definitions of the forward and backward $discreteFouriertransforms$ (DFT) are the following:[1]

$$\hat{X}_k = F\{X_n\}_k = \sum_{n=0}^{N-1} X_n e^{\frac{-2\pi ikn}{N}}, \tag{3}$$

for $k = 0, 1, 2, ..., N-1$ and

$$X_n = F^{-1}\{\hat{X}_k\}_n = \frac{1}{N}\sum_{k=0}^{N-1} \hat{X}_k e^{\frac{2\pi ikn}{N}}, \tag{4}$$

for $n = 0, 1, 2, ..., N-1$.[1] This means that the lower half of the array $\hat{X}_k$ stores modes with positive frequencies, whereas the upper half stores modes with negative frequencies.[1] For real input signals X_n the Fourier coefficients with negative frequencies are just complex conjugates of the coefficients with positive frequencies:[1]

$$\hat{X}_{N-k} = \hat{X}_k^*. \tag{5}$$

The data stored in the upper half of the complex array $\hat{X}_k$ is therefore redundant when the input signal is real.[1]

Quite often we are interested in the distribution of wave amplitudes of the signal across the frequency range.[1] A common measure for the distribution of wave amplitudes is the power spectral density (or power spectrum):[1]

$$P_k = \frac{T^2}{N^2}|\hat{X}_k|^2. \tag{6}$$

The spectrum P_k can reveal interesting features of a signal which are sometimes not obvious when looking in the time domain.[1] The peaks in the spectrum represent the dominant modes which are the most essential for describing the signal.[1] Flat spectra with no apparent peaks correspond to random (white) noise signals.[1]

Another example where the Fourier transform come in handy is the correlation function for signals $X(t)$ and $Y(t)$:[1]

$$C_{X,Y(\tau)} = \lim_{T\to\inf}\frac{1}{T}\int_0^T X(t+\tau)Y(t)^* dt. \tag{7}$$

The correlation function tells us how well do the signals match when shifted by a time lag τ.[1] The correlation function when computer with $X = Y$ is known as the autocorrelation function.[1]

For discrete signals of length N we have to take into account that the number of data points in the overlapping region changes when we vary the discrete time lag l.[1] A first step in the definition of the discrete correlation function requires that we extend X and Y arrays with N zeros:[1]

$$\tilde{X}_n = (X_0, X_1, ..., X_{N-1}, 0, 0, ..., 0), \tag{8}$$

and

$$\tilde{Y}_n = (Y_0, Y_1, ..., Y_{N-1}, 0, 0, ..., 0). \tag{9}$$

a)Electronic mail: blackcubes@ucla.edu.

The size of the extended arrays is now $2N$.[1] Using the new arrays, we may write the discrete correlation as:[1]

$$C_{X,Y}(l) = \frac{1}{N-|l|}\sum_{n=0}^{2N-1} \tilde{X}_{n+l}\tilde{Y}_n^*. \tag{10}$$

The maximal range for the lag l is from $-N+1$ to $N-1$ though the quality of the estimates degrates for large lags because there are fewer points in the region where the signals overlap.[1] The above definition assumes a circular correlation (i.e. periodic data).[1] That is, whenever the array index $n+1$ for $\tilde{X}$ exceeds the arrays bounds it is "wrapped" to become in-bounds using the rule $\tilde{X}_{n\pm 2N} = \tilde{X}_n$.[1]

Using the definitions (3) and (4) for the forward and backward DFT it can be shown that the correlation function can be conveniently written as:[1]

$$C_{X,Y}(l(n)) = \frac{1}{N-|l(n)|}F^{-1}\{F\{\tilde{X}_n\}_k \cdot F\{\tilde{Y}_n\}_k^*\}_n. \tag{11}$$

$C_{X,Y}$ is an array of size $2N$ and the lag l can be written as a function of the array index $n = 0, 1, 2, ..., 2N-1$ as:[1]

$$l(n) = \begin{cases} n, & \text{if } 0 \le n \le N, \\ n-2N, & \text{if } N < n < 2N \end{cases} \tag{12}$$

The lower half of the inverse DFT of $\hat{\tilde{X}}_k\hat{\tilde{X}}_k^*$ corresponds to correlations for positive lags, whereas the upper half of the array corresponds to negative time lags.[1] A division by zero occurs for the element with index N in $C_{X,Y}$ because the normalization faction in Equation (11) is zero.[1] This particular element can be safely discarded in any analysis because it corresponds to a situation where the size of the overlapping region for the X and Y arrays is zero.[1] Instead of computing the correlations from the raw data X_n, Y_n it is common practice to normalize the signals using:[1]

$$X_n \to (X_n - \bar{X})/\sigma_X, \tag{13}$$

$$Y_n \to (Y_n - \bar{Y})/\sigma_Y, \tag{14}$$

where $\bar{X}$, $\bar{Y}$ are the time averages of X_n, Y_n, and σ_X, σ_Y are the standard deviation of X_n, Y_n, respectively.[1] When the signals are normalized as above the correlation function will be bounded between -1 and 1.[1] Values close to 1 indicate very high correlations, whereas values close to -1 indicate high anticorrelations.[1]

For this project, we will perform Fourier analysis on two different data sets: monthly mean sunspot numbers and cosmic ray counts over a period of $T = 50$ years from beginning of 1996 to the end of 2015.[1] The sunspot numbers are provided in the file SUNSPOTS_MONTHLY.TXT and information about the data is given in SUNSPOTS_INFO.TXT.[1] The cosmic ray counts are provided in the file RAYS_MONTHLY.TXT and information about the data is given in RAYS_INFO.TXT.[1] Both files contain $N = 600$ data points.[1] The provided data contains several columns.[1] For the analysis I will use:[1]

- 2nd column of SUNSPOTS_MONTHLY.TXT (monthly mean total sunspot number),
- 3rd column of RAYS_MONTHLY.TXT (correlated cosmic ray count rates).

Sunspots are dark spots on the surface of the Sun with very intense magnetic fields.[1] They typically last for several days, although very large ones may last for weeks.[1] Sunspot measurements have been performed on a regular basis for over 150 years.[1] The number of sunspots is used as a measure of solar (magnetic) activity with a cycle of approximately 11 years.[1]

Cosmic rays are high energy charged particles mainly originating from outside the solar system.[1] They can effect cloud formation through ionization of atmospheric molecules, and they constitute a fraction of the total annual radiation exposure on Earth.[1] During periods of increased solar activity the interplanetary magnetic field strengthens making it more difficult for the cosmic rays to reach the Earth.[1] The cosmic ray flux is therefore anticorrelated with solar activity.[1]

II. RESULTS

A subplot graph of the corrected count rates of the sun rays and the number of monthly mean total of the sunspots is provided on Figure 1, and they would be used for the analysis of this report.

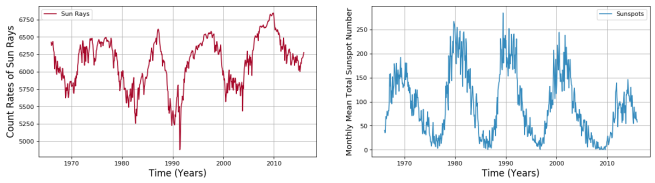

FIG. 1: Graph of corrected count rates of the sun rays and the number of monthly mean total of the sunspots

The power spectra of the sun rays was calculated, and an estimate time period of the dominant nonzero frequency mode in the spectrum is approximately at every 21 years. A subplot graph for both the frequency and period of the sun rays is shown on Figure 2.

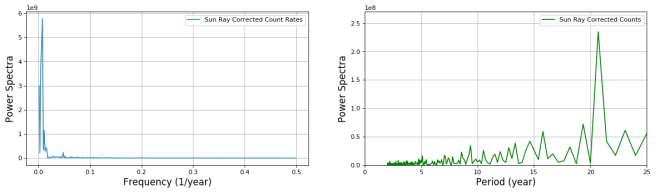

FIG. 2: Power Spectra of the sun rays in both the frequency and period

The power spectra of the sunspots was calculated, and an estimate time period of the dominant nonzero frequency mode in the spectrum is approximately at every 67 years. A subplot for both the frequency and period of the sunspots is shown on Figure 3.

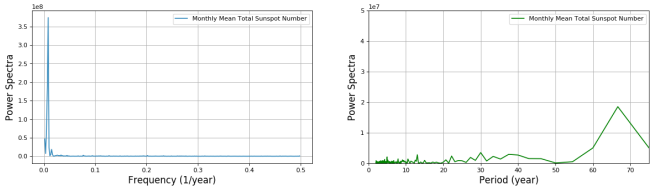

FIG. 3: Power Spectra of the sunspots in both the frequency and period

Instead of using Equation (10) to calculate the autocorrelation and the cross-correlation, I used a different approach such that it is replaced with:

$$\frac{E[(X_k - \bar{X}_k)(Y_k - \bar{Y}_k)]}{\sigma_X \sigma_Y}, \tag{15}$$

where E is the expectation value operator. I also use a "numpy.correlate" to do the autocorrelation and cross-correlation such that it is defined in signal processing. For the input sequences, I used either Equations (13) or (14) and I used "full" for the mode to generate a convolution. The same applies for the cross-correlation where I used both Equations (13) and (14) for the input sequences. With this, I did not have to use the Fourier Transform definition from Equation (10).

The graphs of the normalized autocorrelation functions of the signals of the sunspots and sun rays is provided in Figure 4:

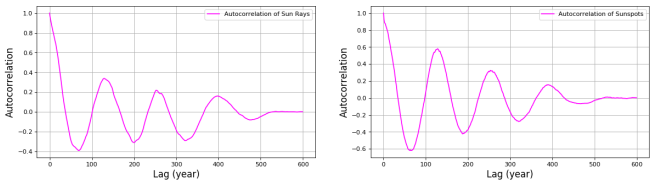

FIG. 4: Autocorrelation of the sunspots and sun rays

On Figure 5 shows a closer view of the autocorrelation for the sun rays where it begins to dip in the first cycle, and the estimated length of the cosmic ray cycle is about 70 years.

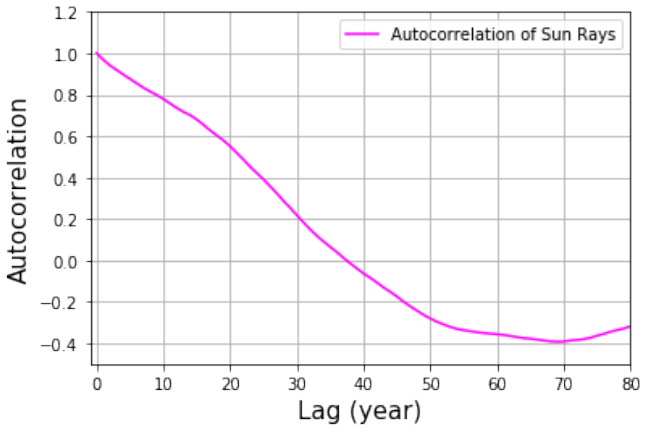

FIG. 5: Closer view of the autocorrelation for the sun rays at a range of $(-0.5, 1.2)$ on the y-axis, and a domain of $[-1, 80]$ on the x-axis

On Figure 6 shows a closer view of the autocorrelation for the sunspots where it begins to dip on the first cycle, and the estimated length of the solar cycle is about 65 years.

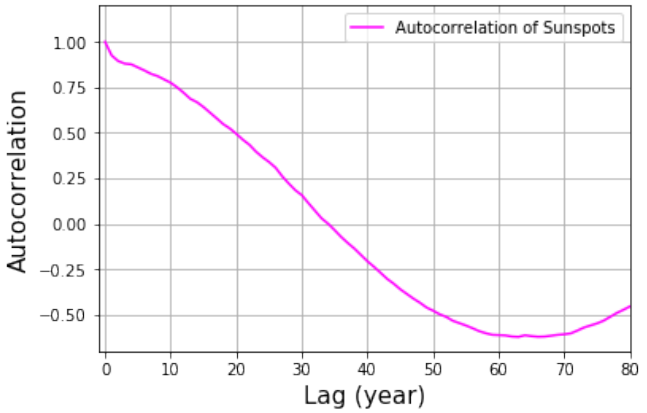

FIG. 6: Closer view of the autocorrelation for the sunspots at a range of $(-0.7, 1.2)$ on the y-axis, and a domain of $[-1, 80]$ on the x-axis

The cross-correlation between the sun rays and sunspots is shown on Figure 7, and a subplot which shows the cross-correlation as well as the autocorrelation of the sunspots and sun rays is shown on Figure 8.

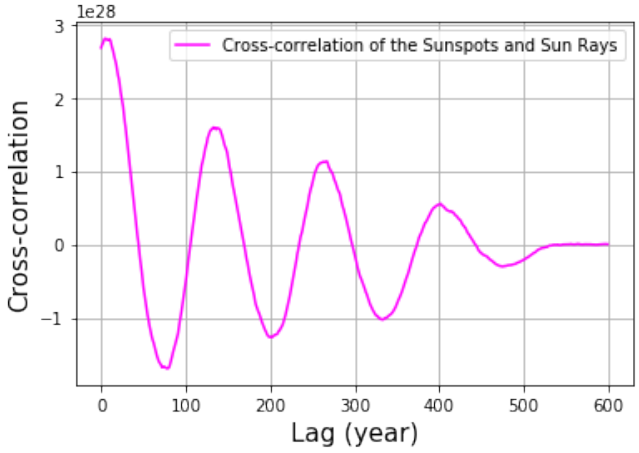

FIG. 7: Cross-correlation between the sun rays and sunspots

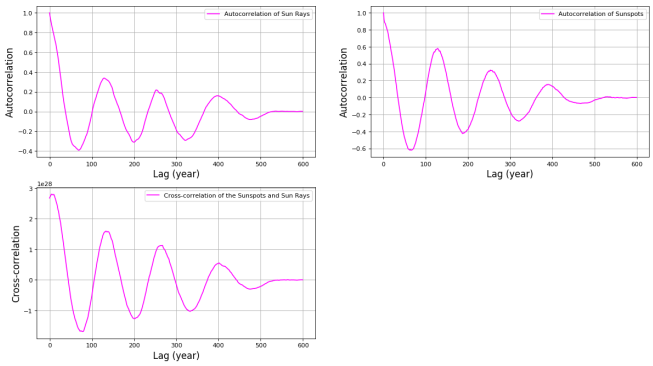

FIG. 8: Three subplots of the cross-correlation between the sun rays and sunspots, autocorrelation of the sunspots, and the autocorrelation of the sun rays

Figure 9 shows a closer view of the cross-correlation:

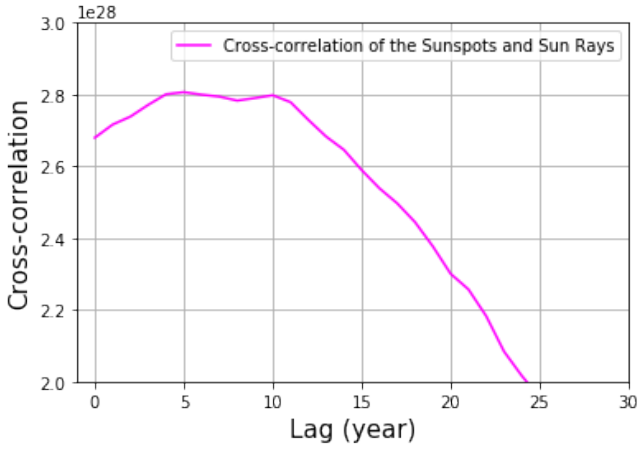

FIG. 9: Power Spectra of the sunspots in both the frequency and period

According to Figure 9, the value of the correlation coefficient $C_{X,Y}(l = 0)$ is at a value of 2.7×10^{28}. The location of the global minimum of the cross-correlation function is when the first inflection point occurs at approximately 2.8×10^{28} of the correlation coefficient and at 7 years on the lag time. This is not at $l = 0$, but shifted slightly to the right for the normalized cross-correlation function.

III. CONCLUSION

By taking the data of the sunspots and sun ray activities, one could see the cycle of each of them. Several calculations were made for these particular data: the calculation of the time period of the dominant nonzero frequency mode in the spectrum for the sun rays and sunspots are 21 years and 67 years, respectively; and the estimated time period of the dominant nonzero frequency mode in the spectrum for cosmic ray cycle and solar cycle are 70 years and 65 years, respectively. Lastly, the correlation coefficient was found at a value of $C_{X,Y}(l = 0)2.7 \times 10^{28}$, and the global minimum shifted to the right where it was not at $l = 0$ was approximately at a correlation coefficient of 2.8×10^{28} at a lag time of 7 years. With the measurements of sunspots and sun ray activities, one could even go as far as to use these data and calculate the greenhouse gases, and whether the planet is affected by global warming.

IV. REFERENCES

[1] J. Samani, .