Journal of Mathematics Research; Vol. 5, No. 1; 2013
ISSN 1916-9795 E-ISSN 1916-9809
Published by Canadian Center of Science and Education

On the Determination of Cycles and Randomness of Sunspot Series Using the Periodogram

Iberedem A. Iwok[1]

[1] Department of Mathematics/Statistics, University of Port Harcourt, Port Harcourt, Rivers State, Nigeria

Correspondence: Iberedem A. Iwok, University of Port Harcourt, P.M.B. 5323, Port Harcourt, Rivers State, Nigeria. E-mail: ibywok@yahoo.com

Received: September 9, 2012 Accepted: January 30, 2013 Online Published: February 19, 2013

doi:10.5539/jmr.v5n1p102 URL: http://dx.doi.org/10.5539/jmr.v5n1p102

Abstract

A periodogram device was applied to the sunspot series X_t^* . The analysis resulted in a period estimate of $10.8 \approx 11$ years. The periodogram was displayed graphically and the largest peak corresponded to a frequency of 0.0926. A significant test on randomness of the series was also carried out to ascertain this result. It was observed that the period 10.8 which accounted for 80.68% of the variance of X_t was statistically significant at the 0.05 level.

Keywords: frequency, period, amplitude, white noise

1. Introduction/Review

Sunspots have been observed since ancient times and is believed to be solar magnetic disturbances manifesting as dark spots on the surface of the sun. In 1610, shortly after viewing the sun with his new telescope, Galileo made the first European observations of sunspots. Continuous daily observations started at the Zurich observatory in 1849 and earlier observations have been used to extend the records back to 1610 (Hathaway, 2010).

Schwabe (1843) collected 17 years of sunspot observations while searching for intramercurial planets. His observations revealed an 11-year periodicity in the number of visible sunspots. However, when Schuster (1906) analysed the sunspot series, the period was found to be 11.125 years.

Rogers and Richards (2004), analysed archival data on sunspot numbers and sunspot areas to derive the solar activity cycles based on these variables. From their work, they were able to identify 11-year Schwabe cycle, Hale cycle, Gleissberg cycle and a cycle at approximately 10 years.

Festus (2010) also confirm the existence of cycles in sunspot numbers using Turkey's smoothing method, and this was calculated to be 10.94 years.

Xu et al. (2010) applied the empirical mode description and autoregressive model to long term sunspot numbers. The method was evaluated using data of the solar 23; and the result was remarkably the predictions made by the solar dyanamo and precursor approaches for cycle 23.

The investigation of the sunspot time series has led to advances in spectrum estimation. As an example, Yule (1927) introduced the concept of a finite parameter model for a stationary random process with special references to wolfer's sunspot numbers.

One way of analyzing time series is based on the assumption that it is made up of sine and cosine waves with different frequencies. The assumption applies the concept of periodogram first introduced by Schuster (1898). Schuster proposed the word as a variable quantity which corresponds to the spectrum of a luminous radiation. He showed that the periodogram could yield information on periodic components of a time series and could be applied even when the periods are not known a priori.

In the olden era, periodogram was used to investigate hidden periodicities and estimate the amplitude of a sine component of known frequency buried in noise. However, this work seeks to extend the importance of the periodogram in identifying which, if any, periodic components explain a large enough percentage of the variance in the time series of interest, and to use the periodogram to test the significance of the result obtained.

From Fourier series model, it could be recalled that a time series of length N can be exactly reproduced by summing $N/2$ sinusoidal wave forms cycle lengths of $N/1, N/2, \ldots, N/(N/2)$ or two observations. The goal of the

periodogram analysis is to partition the sum of squares total (SS_T) for an overall time series of length N into a set of $N/2$ sum of squares (SS) component that correspond to the amount of variance accounted for by each of these cycles.

In addition to the above advantage, if we assume that the time series consist of approximately equal amount of variance due to each of the $N/2$ periodic components, then it will be wise to consider a periodogram as a tool for testing whether a particular series is white noise or not.

2. Method of Analysis

The technique to be used in this work is the periodogram device. From the plot of the raw data (see Figure 1), it would be reasonable to assumed that the time series (X_t^*) is made up of sine and cosine wave form with different frequencies (f_i).

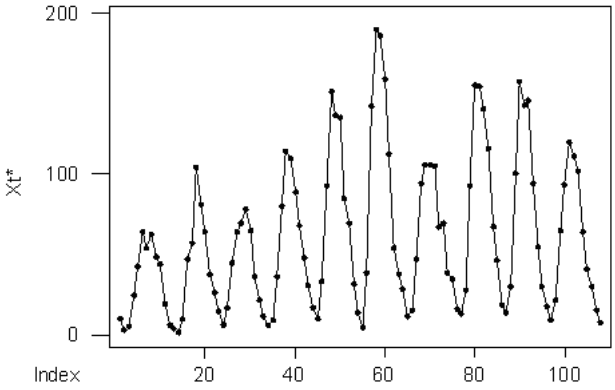

Figure 1. Raw data plot of the sunspot numbers

Data Source: http://www.ngdc.noaa.gov/stp/SOLAR/ftpsunspotnumber.html (1900-2007).

This can be represented using the Fourier series expression:

$$X_t = \mu + \sum_{i=1}^{q} (\alpha_i c_{it} + \beta_i s_{it}) + \varepsilon_t \tag{1}$$

estimated by

$$X_t = \overline{X} + \sum_{i=1}^{q} (a_i c_{it} + b_i s_{it}), \tag{2}$$

where $q = N/2$, $c_{it} = cos2\pi f_i t$, $s_{it} = sin2\pi f_i t$, $a_i = \frac{2}{N}\sum_{t=1}^{N} X_t c_{it}$, $b_i = \frac{2}{N}\sum_{t=1}^{N} X_t s_{it}$, $\varepsilon_t \sim NIID\left(0, \sigma^2\right)$, $\overline{X}$ is the level about which the series fluctuate, $X_t = X_t^* - \overline{X}$, period $= T = N/i$ and $f_i = i/N$ is the i^{th} harmonic of the fundamental frequency $1/N$.

Then, the periodogram $I(f)$ consist of the q values:

$$I(f_i) = \frac{N}{2}\left(a_i^2 + b_i^2\right) \tag{3}$$

It should be noted here that the periodogram is simply the sum of squares associated with the pair of coefficients and the frequencies.

Thus, the proportions of each sum of squares relative to the SS_T is given by:

$$SS_{p(i)} = \frac{SS_{(i)}}{SS_T};$$

that is dividing each SS by SS_T.

In the above expressions, it is assume that the series X_t contains a systematic sine and cosine components with amplitude, phase angle $\emptyset$ and frequency f_i, so that

$$X_t = \mu + \alpha\cos(2\pi f_i t) + \beta\sin(2\pi f_i t) + \varepsilon_t \quad (4)$$

with $Rsin\emptyset = \alpha$ and $Rcos\emptyset = \beta$.

If this assumption were true, then $I(f_i)$ would tend to be inflated because $E[I(f_i)] = 2\sigma^2 + \frac{N(\alpha^2+\beta^2)}{2} = 2\sigma^2 + \frac{NR^2}{2}$ and the hypothesis that X_t is white noise is rejected.

However, if X_t were white noise. That is, truly random, containing no sinusoidal component; then we have: $X_t = \mu + \varepsilon_t$ and $E[I(f_i)] = 2\sigma^2$; $I(f_i) \sim \sigma^2\chi^2_{(2)}$.

3. Data Analysis and Results

The analysis was carried out using the *Minitab* software.

For each of the $N/2 = 108/2 = 54$ cyclic components, a and b coefficients were calculated. From these coefficients, the periodogram ordinates or intensities $I(f_i)$ were obtained as shown in Table 1 below:

Table 1. Periodogram ordinates of sunspot data (1900-2007)

Harmonics i	a_i	b_i	Frequency f_i	Period T_i	Periodogram $I(f_i)$	Proportion
1	-394.00	-17.5370	1/108	108.000	4199676	0.013264
2	36.00	-7.6852	2/108	54.000	36587	0.000116
3	-704.00	-1.4074	3/108	36.000	13381685	0.042262
4	188.0	92.0000	4/108	27.000	1182816	0.003736
5	26.00	7.0800	5/108	21.600	19605	0.000062
6	-132.00	1.9200	6/108	18.000	470548	0.001486
7	-138.00	-3.5800	7/108	15.429	5145534	0.001625
8	-316.00	-0.8400	8/108	13.500	2696131	0.008515
9	-212.00	-8.8200	9/108	12.000	1215588	0.003839
10	-3076.00	11.9400	10/108	10.800	255471801	0.806838
11	746.00	10.3400	11/108	9.818	15028819	0.047464
12	144.00	-20.0111	12/108	9.000	570684	0.001802
13	870.00	-3.3000	13/108	8.308	20436594	0.064543
14	-224.00	2.1200	14/108	7.714	1354873	0.004279
15	28.00	-1.2800	15/108	7.200	21212	0.000067
16	30.00	-0.3200	16/108	6.750	24303	0.000077
17	2.90	-0.6400	17/108	6.352	238	0.000001
18	8.94	2.7000	18/108	6.000	2357	0.000007
19	2.44	-5.9000	19/108	5.684	1101	0.000003
20	4.04	2.0700	20/108	5.400	637	0.000002
21	-6.12	-2.3000	21/108	5.143	1154	0.000004
22	0.15	3.4600	22/108	4.909	324	0.000001
23	1.41	-1.3000	23/108	4.696	99	0.000000
24	1.26	0.4800	24/108	4.500	49	0.000000
25	0.42	1.3600	25/108	4.320	55	0.000000
26	-0.92	0.8800	26/108	4.154	44	0.000000
27	-0.41	-1.8600	27/108	4.000	98	0.000000
28	-0.72	0.9400	28/108	3.857	38	0.000000
29	0.18	-0.1400	29/108	3.724	1	0.000000
30	1.08	-1.0200	30/108	3.600	60	0.000000
31	-2.56	-0.4126	31/108	3.484	181	0.000001
32	-0.17	0.8200	32/108	3.375	19	0.000001
33	-2.44	1.9800	33/108	3.272	267	0.000000
34	2.35	-0.7800	34/108	3.176	166	0.000001
35	0.24	-0.4600	35/108	3.086	7	0.000001
36	0.19	0.5400	36/108	3.000	9	0.000000
37	-0.65	0.6800	37/108	2.919	24	0.000000
38	-0.80	-0.5200	38/108	2.842	23	0.000000
39	-0.93	0.7200	39/108	2.769	37	0.000000

40	-0.15	0.9600	40/108	2.700	25	0.000000
41	1.69	0.0600	41/108	2.634	77	0.000000
42	0.17	0.6000	42/108	2.571	10	0.000000
43	1.12	0.6600	43/108	2.512	46	0.000001
44	-1.57	1.6600	44/108	2.455	141	0.000000
45	0.44	-2.5600	45/108	2.400	182	0.000000
46	1.26	0.4400	46/108	2.348	48	0.000000
47	0.57	-1.1200	47/108	2.298	43	0.000000
48	0.19	-0.1200	48/108	2.250	1	0.000000
49	-0.78	-1.1200	49/108	2.204	50	0.000000
50	-2.09	0.0600	50/108	2.160	118	0.000000
51	0.70	-0.4200	51/108	2.118	18	0.000000
52	0.46	-0.9600	52/108	2.077	31	0.000000
53	0.39	-1.9200	53/108	2.038	104	0.000000
54	1.15	0.0006	54/108	2.000	36	0.000000
					SS_T = 316633374	

Data Source: http://www.ngdc.noaa.gov/stp/SOLAR/ftpsunspotnumber.html (1900-2007).

From the above table, it is seen that the only periodogram component that has a large SS and therefore accounts for a relatively large proportion of the variance in the time series corresponds to a frequency of $10/108 = 0.0926$ or a cycle length of 10.800 years. The SS proportion ($SS_{p(i)}$) of this outstanding component with the largest SS is 0.80684. This means that the 10.800 cycle account for 80.684% of the variance of X_t.

4. The Peridogram Plot

The plot of the periodogram ordinates (SS) versus frequency (f_i) is displayed below:

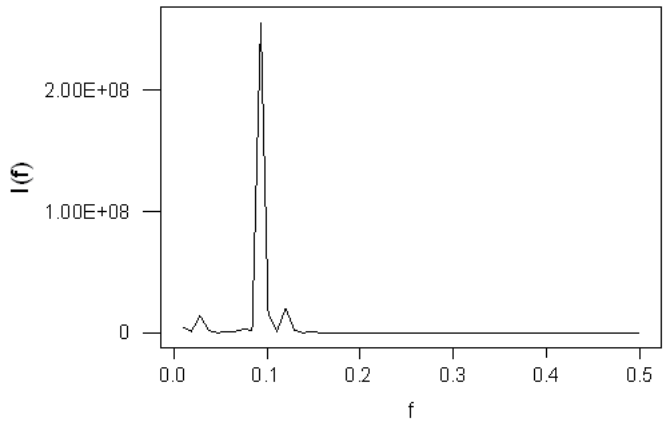

Figure 2. The periodogram plot of the sunspot numbers (1900-2007)

Visual examination of this plot shows that the largest peak occurs at a frequency of 0.0926 which corresponds to a period of 10.8 years.

However, large peaks can arise by chance; so it is desirable to use a statistical significance test to assess the result.

5. The Significance Test

For this analysis, the null hypothesis H_0 is that X_t is white noise. Of course, if H_0 were true, we would expect approximately equal SS values across the entire periodogram. In this work, however, the test developed by Fisher (1929) and critical values tabled by Russell (1985) is used. This test is preferred because it takes into account the inflated risk of type I error (α) that arises when many periodic components are examined.

Let SS_L be the largest intensity estimate, then the test statistic is

$$F^* = \frac{SS_L}{SS_T} = \frac{255471801}{316633374} = 0.80684$$

This value when compared with the critical value of 0.12334 shows that the largest peak of period 10.8 is statistically significant and so H_0 is rejected. Similarly, the second largest SS value (15028819) of period 9.818 was also tested, and was shown to be insignificant. This eventually terminates the test procedure. Therefore, the only significant periodic component of the sunspot series is 10.8 years.

6. Conclusion

Recently, there are strong indications that the warming and cooling of the earth might be due to the changes in the number of observed sunspots (Eyeni, 2010). In order to provide solutions to the global warming in particular, several scientists have engaged in determining the cycle of sunspot numbers which is one of the factors affecting the warming and cooling nature of the earth. Several periods have been identified by various scientists using different methods without undergoing test to ascertain their claims. This work, however, has analysed the sunspot series using the periodogram device. The period has been identified to be 10.8 years with a frequency of 0.0926. This value is not only close to the values obtained by the afformentioned scientists but is supported by the significant test.

References

Box, G. E. P., & Jenkins, G. M. (1970). Time series analysis, forecasting and control. Holden day. San Francisco. New York.

Eyeni, C. H. (2010). Factors affecting the warming and cooling of the earth. *International Journal of Geography, 3*(2), 8-13.

Festus, H. K. (2010). Analysis of sunspot numbers using Turkeys method. *International Journal of Physical Science and Creativity, 3*(4), 13-18.

Hathaway, D. H. (2010). Update of sunspot cycle. National aeronautics and space administration. Marshall flight space centre.

Rogers, M. L., & Richards, M. T. (2004). Periodogram analysis of sunspot numbers and areas. *Bulletin of the American Astronomical Society, 8*(2).

Schuster, S. (1898). On the investigation of hidden periodicities. *Terr. Mag., 3*(13), 42-48.

Schuster, A. (1906). On the periodicities of sunspots. *Philosophical Transactions of the Royal Society, 206*(402-412), 69-100. http://dx.doi.org/10.1098/rsta.1906.0016

Schwabe, S. H. (1843). Solar observations during 1843. *Astronomische Nachrichten, 20*(495), 234-235.

Yule, G. U. (1927). On the method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society, 226*, 267-298. http://dx.doi.org/10.1098/rsta.1927.0007

Xu, T., Wu, J., Wu, Z. S., & Li, Q. (2010). Long term sunspot number prediction based on EMD analysis and AR model. *Chinese Journal of Astronomy and Astrophysics, 8*(337).