Solar Phys (2009) 257: 363–369
DOI 10.1007/s11207-009-9380-3

Evidence of 6 000-Year Periodicity in Reconstructed Sunspot Numbers

M.A. Xapsos · E.A. Burke

Received: 23 December 2008 / Accepted: 7 May 2009 / Published online: 27 May 2009

Abstract Reconstructed sunspot data are available that extend solar activity back to 11 360 years before the present. We have examined these data using Hurst analysis, a moving average filter, and Fourier analysis. All of the procedures indicate the presence of a long term (≈6 000 year) cycle not previously reported. A number of shorter cycles formerly identified in the literature by using Fourier analysis, Bayes methods, and maximum entropy methods were also detected in the reconstructed sunspot data.

Keywords Solar activity, periodic · Hurst analysis · Smoothing filter · Fourier analysis

1. Introduction

Solanki *et al.* (2004) reported reconstructed sunspot numbers extending back in time almost 11 400 years. The reconstruction was based upon ^{14}C in tree rings and ^{10}Be in ice cores. Past studies have shown that long-term records of many natural phenomena exhibit persistent behavior. These include floods, rainfall, temperatures, and sunspot numbers (SSN). A classic example is the Nile River, where prolonged periods of dryness were followed by periods of floods. A method of analyzing this behavior, now known as rescaled range analysis (commonly referred to as R/S analysis), was developed by the hydrologist Hurst (1951) and Hurst, Black, and Simaika (1965). An important element in this analysis involves calculating the cumulative deviation from the average of a stochastic quantity (in the present case SSN) over the time period of interest. Persistence is characterized by a parameter now commonly referred to as the H coefficient. Coefficients with a value significantly greater than 0.5 are taken to indicate correlated or persistent behavior.

In a paper by Ruzmaikin, Feynman, and Robinson (1994), R/S analysis was applied to ^{14}C data extending over a period of 3 000 years and the authors derived an H coefficient of

M.A. Xapsos (✉)
NASA Goddard Space Flight Center, Greenbelt, MD, USA
e-mail: Michael.A.Xapsos@nasa.gov

E.A. Burke
Consultant, Woburn, MA, USA

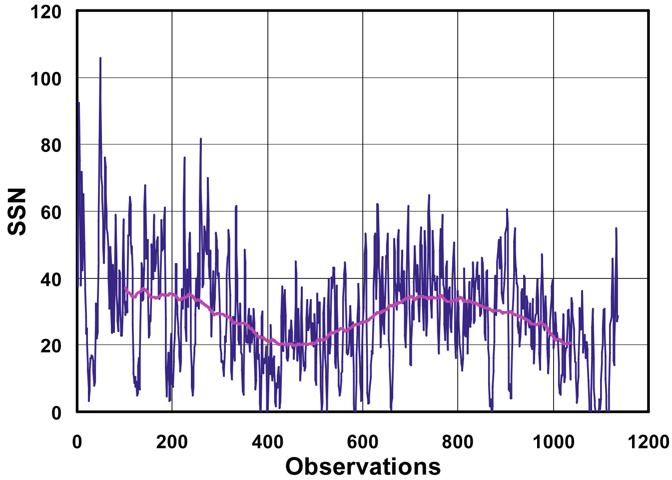

Figure 1 Reconstructed sunspot record from Solanki *et al.* (2004). Also shown are results for a moving average filter applied to the data as described in a later section. The time period covers 11 360 years with observations in 10-year increments.

0.84. In this paper the Hurst method has been applied to the 11 360 years of data reported by Solanki *et al.* (2004). The value of H derived from this longer series is 0.81, which is quite close to the 0.84 value reported by Ruzmaikin, Feynman, and Robinson (1994) given the differences in the data sets. However, the focus in this paper is not the Hurst coefficient but what appears to be evidence of a long cycle with a period of about 6 000 years uncovered during the course of R/S analysis.

2. Hurst Analysis

In the following description the notation of Beran (1998) is used. For a period of T years beginning at time t the cumulative measure of SSN activity is defined as

$$Y_{t+T} = \sum_{i=t}^{t+T} X_i. \tag{1}$$

In this application the X_i are taken as the 10-year averaged SSN in the Solanki data set (Solanki *et al.*, 2004). Y_{t+T} is the sum of the 1 136 SSN that span a period of $T = 11\,360$ years. The cumulative deviation is then

$$\Delta Y_{t+T} = \sum_{i=t}^{t+T} (X_i - \bar{Y}_{t+T}), \tag{2}$$

where $\bar{Y}_{t+T}$ is the average of the stochastic quantity X_i. The cumulative deviation represents the difference between the cumulative measure of SSN at a given time and a cumulative calculation based on the average over the total time period of interest.

Hurst analysis provides a method for determining whether or not a stochastic series exhibits long-range correlation. The series of N observations are divided into segments of

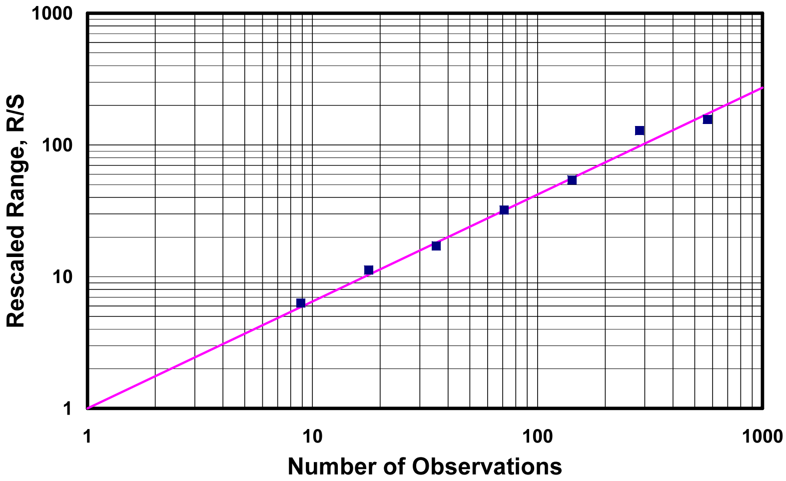

Figure 2 Hurst fit to the data of Solanki *et al.* (2004) resulting in a Hurst coefficient of $H = 0.81$.

$N/2$,$N/4$, $N/8$, etc. The difference between the maximum and minimum values in each segment, R, is then divided by the standard deviation, S, of the elements in that same segment. For cases of multiple segments of the same size (*e.g.*, eight elements of length $N/8$) the results are averaged to obtain the R/S value shown in Figure 2. The procedure is repeated and at each stage the values are summed and averaged. When the averaged values are plotted versus the number of observations in the associated population on a log – log graph they are typically found to fall on a straight line as shown in Figure 2.

Hurst found that most long-term natural phenomena were approximated by the expression

$$\frac{R}{S} = \left(\frac{N}{2}\right)^H, \tag{3}$$

where N is the number of observations. For a correlated Gaussian random series, the power index H is substantially greater than 0.5. It approaches a maximum value of 1.0 as the correlation or persistence increases.

A power index of about 0.5 is taken as an indication of uncorrelated data and values of the power index greater than 0.7 strongly indicate long-term correlation. The results obtained here clearly support the fact that long-term correlation exists in the data of Solanki *et al.* (2004).

A first step in Hurst analysis involves calculation of the cumulative deviation of SSN (net SSN) over the period of interest using Equation (2). As shown in Figure 3, when applied to the total record compiled by Solanki *et al.* (2004) two periods on the order of thousands of years are revealed during which solar activity markedly exceeded the average for the period of recorded data. It should be noted, however, that this analysis may not yield an accurate reading of the cyclical period because the end points in Figure 3 are forced to zero by Equation (2). Nonetheless, it appears there are two cycles. Although they could be ascribed to chance, that alternative seems improbable. Simulations of random processes typically do not show periodic patterns of the type observed in Figure 3. Usually, the simulations have a significant portion of their paths on negative values of the ordinate.

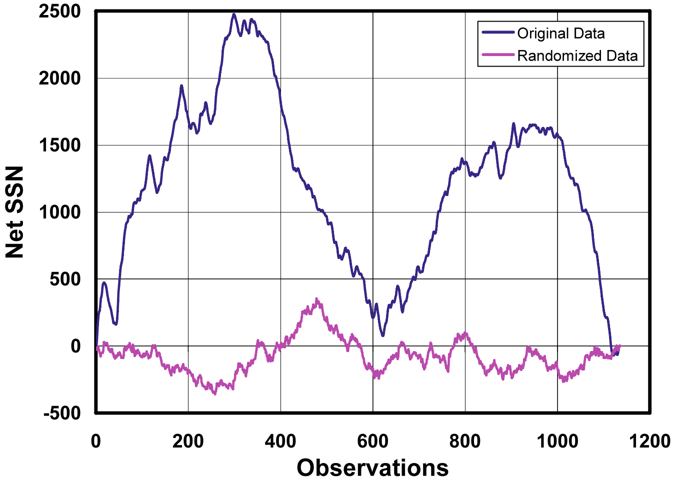

Figure 3 Cumulative deviation of the original and randomized data.

Also shown in Figure 3 is the cumulative deviation analysis of the same set of data treating it as a random distribution at all times. This destroys the correlation in the original data series. Although apparent cycles can arise in stochastic data (see Yule, 1927; Slutzky, 1937) the pattern observed in Figure 3 is unusual.

3. Application of a Smoothing Filter

To view this periodic behavior more directly, a centered moving average filter was applied to the data of Solanki *et al.* (2004) using a window of 201 data points. The result is shown in Figure 1 compared to the original data. The centered moving average that was used can be written

$$\overline{X}_n(i) = \frac{1}{n} \sum_{k=-[(n+1)/2]+1}^{[(n-1)/2]} X(i+k). \tag{4}$$

$\overline{X}_n(i)$is the moving average value at point i, with the average taken over n data points.

The result shown in Figure 1 indicates about one and one-half cycles of a sinusoid. This is consistent with the so-called sinusoidal limit described by Slutzky (1937). According to Kendall and Stuart (1966) there should be a limit described by the difference equation for a sine curve given by

$$x_{i+2} - 2\rho_1 x_{i+1} + x_i = 0. \tag{5}$$

Equation (5) suggests that the sinusoidal data in Figure 1 should yield a constant equal to $2\rho_1$. The average value obtained from the data was 2.00, implying a value for ρ_1 of about 1.0. A least squares fit of the sinusoid establishes a period of 608.5 observations or 6 085 years, given the average of 10 years per observation. The coefficient of correlation was 0.983.

4. Fourier Analysis

A conventional way of detecting cycles in data involves the application of the discrete Fourier transform (DFT). The term periodogram has been widely used in the literature and is often associated with different mathematical expressions. Here we employ the formalism used by Fuller (1996).

The intensity as a function of frequency can be expressed as

$$I_n(\omega_k) = \frac{2}{n}\left[\left(\sum_{t=1}^{n} X_t \cos\omega_k t\right)^2 + \left(\sum_{t=1}^{n} X_t \sin\omega_k t\right)^2\right], \quad k = 0, 1, \ldots, m, \tag{6}$$

where the X_t are the n original observations of the time series and ω_k is the frequency,

$$\omega_k = \frac{2\pi k}{n}. \tag{7}$$

The set of frequencies defined by Equation (7) is often referred to as the Fourier frequencies.

Application of the DFT to the original time series data yields the periodogram defined by Equation (6). As is typical the result is a series of peaks and is shown in Figure 4.

The spectrum derived actually extends to a frequency of 0.5 but all significant peaks occur within the range shown. The significant new result here is the very prominent peak that occurs at a frequency of 0.00176, equivalent to a period of 5 680 years. Previous studies of solar periodicities are based on shorter time series data and do not reflect this ≈6 000-year period. The next largest peak observed is the ≈2 300-year cycle, which has been reported a number of times (see Damon and Sonett, 1991).

The problem of determining whether peaks in a periodogram are significant cyclic components or simply an artifact of noise in the data is central. The tests applied here are based on analysis given in Fuller (1996) and Warner (1998) as well as analytical forms provided by Grenader and Rosenblatt (1957) and Hannan (1967). A statistic that can be used to test

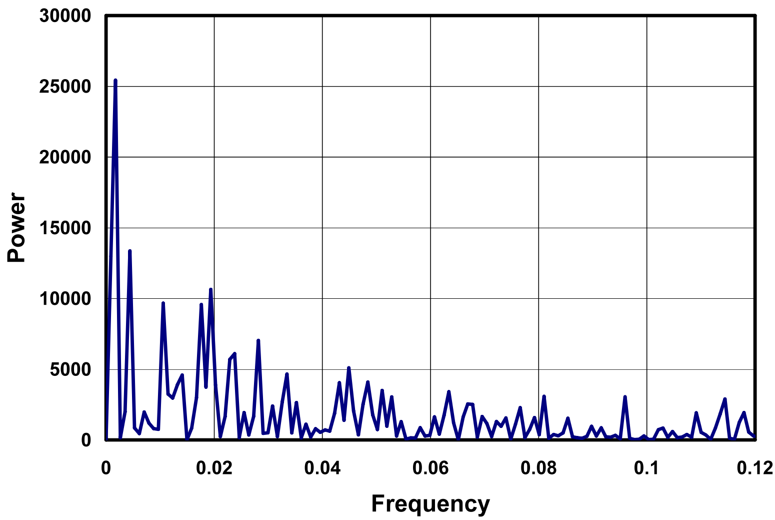

Figure 4 Periodogram of the data of Solanki *et al.* (2004).

Table 1 Critical and observed ratios for the five largest peaks.

Rank	Period (years)	Critical ratio	Observed ratio
1	5 680	10.85	50.72
2	2 272	8.15	26.60
3	516	7.06	21.15
4	947	6.41	19.17
5	568	5.95	18.95

the hypothesis that the largest peak is generated by the random sample is

$$\xi_m = \left(\frac{1}{m} \sum_{k=1}^{m} I_n(\omega_k) \right)^{-1} I_n(L), \qquad (8)$$

where $I_n(L)$ is the largest peak in a sample of m periodogram ordinates. This is simply the intensity of the maximum peak divided by the average of the m peaks in the sample and is referred to as the critical ratio. The critical ratios given by Fuller (1996) vary from 1.99 to 11.454 for values of m ranging from 2 to 1 000. For these data shown in Figure 4, m equals 568 or half of the 1136 observations in the time series. An observed peak with a ratio calculated using Equation (8) is considered significant if the observed ratio exceeds the critical value. Tables of critical values as given by Warner (1998) permit similar tests to be carried out on the five highest peaks in the periodogram.

When these tests are applied to the first five peaks extracted from the data of Solanki *et al.* (2004), the results in Table 1 are obtained. The ranks are based upon the power in the respective peak. The table shows the period associated with the peak, the critical ratio (see Fuller, 1996; Warner, 1998), and our observed ratio from application of Equation (8) to the data.

All of the observed ratios in Table 1 exceed the critical values by a significant amount. The peaks ranked from 2 to 5 have previously been reported (see Damon and Sonett, 1991; Vasiliev and Dergachev, 2002; Dergachev, 2004). The new result is that ranked number 1 with a period close to 6000 years. The critical ratios above would be exceeded only 1% of the time or less if the peaks were due to a random process. That taken together with the fact that the periodogram in Figure 4 showed other prominent peaks previously reported and detected using different methods of analysis, as shown in Table 2, supports the presence of a long-term (≈6 000-year) cycle.

Table 2 compares our Fourier results with data from Damon and Sonett (1991), Vasiliev and Dergachev (2002), and Dergachev (2004). In the compilation of Damon and Sonett (1991) several different methods of analysis were used, including DFT, Bayes analysis, and maximum entropy.

5. Conclusions

Solanki *et al.* (2004) have provided reconstructed sunspot data back to 11 360 years before the present. We have examined these data using Hurst analysis, a moving average filter, and Fourier analysis. All of the evidence appears to indicate the presence of a long-term ≈6 000-year cycle that has not been reported previously. Tests of the frequencies derived from the discrete Fourier transform indicate that they are significant.

Table 2 Cycles reported in the literature.

Rank	Frequency		Period (years)			
	This work		Vasiliev and Dergachev (2002); Dergachev (2004)	Damon and Sonett (1991)[a]	Damon and Sonett (1991)[b]	Damon and Sonett (1991)[c]
1	0.00176	5 680	–	–	–	–
2	0.00440	2 272	2 400	2 272	2 385	2 241
3	0.01937	516	505	–	512	504
4	0.01056	947	940	909	955	805
5	0.01761	568	570	649	–	–
6	0.02817	355	360	–	353	385
7	0.02377	421	420	–	440	427
8	0.04489	223	–	207	208	208
9	0.03345	299	–	–	–	299
10	0.04313	232	230	–	–	238

[a]DFT.

[b]Bayes analysis.

[c]Maximum entropy.

The apparent presence of a 6 000-year periodicity in the data raises the possibility that evidence of such a cycle has been reported in proxy data such as that for ice cores and cosmogenic radionuclides. Studies of 110 000-year glaciochemical series have revealed indications of a 6 100-year event (see Mayewski *et al.*, 1997).

References

Beran, J.: 1998, *Statistics for Long-Term Memory Processes*, Chapman & Hall/CRC, London/Boca Raton.

Damon, P.E., Sonett, C.P.: 1991, In: Sonett, C.P., Giampapa, M.S., Matthews, M.S. (eds.) *The Sun in Time*, University of Arizona Press, Tucson.

Dergachev, V.A., 2004, In: Stepanov, A.V., Benevolenskaya, E.E., Kosovichev, A.G. (eds.) *Multi-Wavelength Investigations of Solar Activity, IAU Symp.* **223**, 699.

Fuller, W.A.: 1996, *Introduction to Statistical Time Series*, 2nd edn., Wiley, New York.

Grenader, U., Rosenblatt, M.: 1957, *Statistical Analysis of Stationary Time Series*, Wiley, New York.

Hannan, E.J.: 1967, *Time Series Analysis*, Methuen, London.

Hurst, H.E.: 1951, *Trans. Am. Soc. Civ. Eng.* **116**, 770.

Hurst, H.E., Black, R.P., Simaika, Y.M.: 1965, *Long Term Storage: An Experimental Study*, Constable, London.

Kendall, M.G., Stuart, A.: 1966, *The Advanced Theory of Statistics*, Vol. 3, Hafner, New York.

Mayewski, P.A., Meeker, L.D., Twickler, M.S., Whitlow, S., Yang, Q., Lyons, W.B., Prentice, M.: 1997, *J. Geophys. Res.* **102**, 26345.

Ruzmaikin, A., Feynman, J., Robinson, P.: 1994, *Solar Phys.* **149**, 395.

Slutzky, E.: 1937, *Econometrica* **5**, 105.

Solanki, S.K., Usoskin, I.G., Kromer, B., Schüssler, M., Beer, J.: 2004, *Nature* **431**, 1084.

Vasiliev, S.S., Dergachev, V.A.: 2002, *Ann. Geophys.* **20**, 115.

Warner, R.M.: 1998, *Spectral Analysis of Time-Series Data*, Guilford, New York.

Yule, G.U.: 1927, *Philos. Trans. Roy. Soc. Lond. A* **226**, 267.