The sunspot cycle revisited

Nick Lomb
Sydney Observatory, PO Box K346, Haymarket NSW 2000, Australia
Email: nickl@phm.gov.au

Abstract. The set of sunspot numbers observed since the invention of the telescope is one of the most studied time series in astronomy and yet it is also one of the most complex. Fourteen frequencies are found in the yearly mean sunspot numbers from 1700 to 2011 using the Lomb-Scargle periodogram and prewhitening. All of the frequencies corresponding to shorter term periods can be matched with simple algebraic combinations of the frequency of the main 11-year period and the frequencies of the longer term periods in the periodogram. This is exactly what can be expected from amplitude and phase modulation of an 11.12-year periodicity by longer term variations. Similar, though not identical, results are obtained after correcting the sunspot number series as proposed by Svalgaard. On looking separately at the amplitude and phase modulation a clear relationship is found between the two modulations although this relationship has broken down for the last four solar cycles. The phase modulation implies that there is a definite underlying period for the solar cycle. Such a clock mechanism does seem to be a possibility in models of the solar dynamo incorporating a conveyor-belt-like meridional circulation between high polar latitudes and the equator.

1. Introduction

Sunspots are an indication of solar activity. Records of sunspot numbers over hundreds of years can be studied to examine the variability of the Sun. These studies can give clues to the nature of the mechanism driving the solar cycle and provide constraints that must be met by any viable theory.

A quick look at a plot of the sunspot number time series indicates a quasi-regular variation of about 11 years that changes in amplitude and length. This 11-year period can largely disappear during great minima such as the Maunder Minimum, which persisted for much of the 17th century.

An obvious feature of the sunspot number time series, one that is so obvious that it is usually missed, is that the definition of the series constrains it to be always positive. Hence, and unlike most other time series with which scientists are familiar, the fluctuation in sunspot numbers is not about zero but about the mean value. Stated mathematically the series is not in the form

\[ A(t) \sin(2\pi ft + P(t)) \]

where \( A(t) \) and \( P(t) \) represent amplitude and phase modulation terms, respectively, but in the form

\[ A(t) \left[ 1 + \sin(2\pi ft + P(t)) \right] \] (1)

As was shown in a previous paper [1], to be referred to as paper 1, this peculiar, always positive, nature of the sunspot time series means that any modulation period must show in the spectrum both as a sidelobe to the main period and as a separate periodicity.
In paper 1 the yearly mean Wolf Sunspot Numbers [2] from 1700 to 1964 were studied by frequency analysis using least squares or LS frequency analysis [3], which has since then been widely used under the name Lomb-Scargle Periodogram [4]. In the paper a set of 14 statistically significant periodicities was found in the frequency spectrum. These periodicities could be described in terms of just three main periods. This was interpreted as indicating that the sunspot series results from the amplitude and phase modulation of the main 11-year period by two longer term periodicities. In addition, a correlation was found between the amplitude and phase modulation in the sense that the stronger the peak of a cycle the earlier it occurred.

The current paper aims to extend the analysis of paper 1 by including the most recent solar cycles and to check whether the sunspot numbers still show the same structure with more recent data. It also aims to better interpret the results on the basis of modern models of the solar dynamo. Additionally, the analysis is to be repeated after increasing the pre-1945 sunspot numbers by 21% following Svalgaard [5] who has found that such corrections are necessary to maintain the uniformity of the series.

2. Analysis
In this paper the yearly mean Wolf sunspot numbers are again used, but this time from 1700 to 2011 with the data taken from the Solar Influences Data Analysis Center (http://sidc.oma.be/sunspot-data/).

2.1. Frequency analysis
The frequency analysis was carried out with the Lomb-Scargle Periodogram using the Peranso period analysis software (http://www.peranso.com/). After each frequency was found the data was prewhitened and the calculation repeated to obtain the next frequency.

![Periodograms](image)

**Figure 1.** Periodograms of the yearly mean sunspot numbers, 1700-2011: the periodogram of the original data (blue), the periodogram after prewhitening by the main 11-year periodicity (red) and the periodogram after prewhitening by all 14 periods found in the data (green). Power on the y axis represents the reduction in the sum of squares due to each frequency with the sunspot numbers scaled to vary from +1 to -1.

The frequency analysis was stopped after finding 14 periods in the data. Figure 1 shows the frequency spectrum of the raw sunspot series together with the spectra after the removal of the main period and after the removal of all 14 periods. As can be seen from the figure, prewhitening is important as the mutual influence of adjacent frequency peaks can shift the position of the peaks and hence the determined frequency.
After the removal of all found periods a Monte-Carlo analysis indicated that the next remaining peak had a false alarm probability of only 0.045±0.015. There are thus probably further statistically significant periodicities remaining in the data, however, they are unlikely to make much contribution to our understanding of its structure. Table 1 lists the periods/frequencies that were found together with their uncertainties. The Peranso program calculates uncertainties with a method due to Schwarzenberg-Czerny [6].

| Observed period (yr) | Observed frequency (cycles/yr) | Predicted frequency (cycles/yr) | Identification |
|----------------------|-------------------------------|---------------------------------|----------------|
| 11.00±0.06           | 0.0909±0.0005                 | 0.0899                          | f_1            |
| 10.03±0.05           | 0.0997±0.0005                 |                                 | f_1+f_01       |
| 10.61±0.08           | 0.0942±0.0007                 | 0.0944                          | f_1+2f_02      |
| 102.07±7.83          | 0.0098±0.0007                 |                                 | f_01           |
| 11.86±0.14           | 0.0843±0.0010                 | 0.0833                          | f_1-3f_02      |
| 449.24±94.61         | 0.0022±0.0005                 |                                 | f_02           |
| 8.46±0.05            | 0.1182±0.0007                 | 0.1193                          | f_1+3f_01      |
| 13.08±0.17           | 0.0764±0.0010                 | 0.0743                          | f_1f_03        |
| 8.12±0.08            | 0.1232±0.0012                 | 0.1213                          | f_1+2f_03      |
| 63.80±5.02           | 0.0157±0.0012                 |                                 | f_03           |
| 9.40±0.09            | 0.1064±0.0010                 | 0.1056                          | f_1+f_03       |
| 43.56±1.42           | 0.0230±0.0007                 |                                 | f_04           |
| 10.34±0.11           | 0.0967±0.0010                 | 0.0966                          | f_1+3f_02      |
| 28.55±1.02           | 0.0350±0.0012                 |                                 | f_05           |

As in paper 1 the shorter periods can all be matched by simple algebraic combinations of the frequency of the main 11-year period and the frequencies of the longer term periods in the periodogram. This is exactly what can be expected when one periodicity of the form in equation 1 is amplitude and phase modulated by longer periods. The predicted value of f_1 was taken from the determined value of f_1+f_01 as paper 1 showed that term was determined with smaller errors.

The analysis was repeated with the sunspot numbers modified following the scheme proposed by Svalgaard [5]. Results similar to those with the unmodified data are obtained though, as expected, with changes to the longer term periods. The longer term periods equivalent to f_01 to f_05 in table 1 become, 99.79, 67.32, 54.19, 170.49 and 44.31 years. The long trend period f_02 from table 1 is now gone, but the period around 100 years remains with the modulation by this period obvious in a visual examination of a plot of the modified sunspot number data. Most interestingly, all five of these longer-term periods in the modified sunspot data are subharmonics of the main 11-year periodicity.

2.2. Amplitude and phase modulation

Various authors have looked at the modulation in phase, generally interpreted as variation in frequency, using highly sophisticated modern time series methods. For example, Kolláth and Oláh [7] follow the frequency variation of the 11-year period, its first harmonic and that of a longer period using the pseudo-Wigner distribution from sound processing. Here, instead, we looked at the modulation in more detail using the conceptually simple technique of fitting by least squares a sine
wave with an 11-year period and a constant for each of the 28 minimum-to-minimum cycles since 1700. The original unmodified yearly mean Wolf sunspot numbers were used. The longer period modulation of the sunspot cycle can be seen in both the amplitudes and the phases as shown in figure 2. As expected from (1), there is a good fit between the constants and amplitudes. As well there is also a clear relation between the phases and the amplitudes. To explore this relationship a linear regression model was fitted between the phases and the amplitudes together with a time term to allow for differences in period from the assumed 11 years.

The regression coefficient for the period of $-0.33\pm0.12^\circ$ per sunspot yields a frequency of 0.0900 cycles per year or a period of 11.11 years. This value for the main solar cycle is in good agreement with the predicted value listed in table 1 that equates to 11.12 years.

As shown in figure 2, there is a clear relationship between the amplitude and the phase modulation, though with occasional deviations as in the late 18th century. Interestingly, the relationship seems to have completely broken down for the last four solar cycles. This relationship is connected to the Waldmeier effect [8], according to which there is an inverse correlation between the rise time of a cycle and its amplitude.

![Figure 2](image.png)

**Figure 2.** Observed (blue diamonds) and predicted phases (red squares) of the 11-year period calculated from a regression fit of the amplitude variations to the phases.

3. **Discussion**

This study has found evidence for long-term modulation of the solar cycle over periods of 28 to 450 years. At this stage there is no indication if any of the modulation periodicities are real or if they just represent a Fourier fit to a random variation. It can be noted though that a period of around 100 years, which matches the well-known Gleissberg cycle [9], does seem to be persistent in the data.

This study confirms the structure of the sunspot time series demonstrated in paper 1: a stable 11-year periodicity (the Schwabe cycle) that is amplitude and phase modulated by the long-term periodicities discussed above. This clearly implies that a clock mechanism must exist within the Sun for the 11-year periodicity to persist in the solar data as was first suggested by Dicke [10].

Modern theories provide a possible clock mechanism in the conveyor-belt-like meridional circulation between high polar latitudes and the equator. In the Babcock-Leighton models of Charbonneau and Dikpati [11] and [12] remnant magnetic flux from decaying sunspots is transported away from the equator by meridional circulation towards the poles generating the poloidal field of the following cycle. This field is transported to the base of the convection zone where shearing by differential rotation leads to a new toroidal field at low latitudes. Buoyant flux tubes rise to the surface as sunspots. In this way the meridional circulation provides the clock regulating the 11-year cycle and maintaining its continuity.
The models predict fluctuations in the meridional circulation and these fluctuations have been observed with the Michelson Doppler Imager on board the SOHO spacecraft [13]. However, according to Charbonneau and Dikpati [11] the meridional circulation can still act as a clock for most of their stimulations exhibit ‘good phase locking, in the sense that their cycle periods rarely depart for more than a few consecutive cycles from their average value’. Moreover, the models also reproduce the amplitude-duration anticorrelation that is related to the amplitude-phase relationship discussed in this paper.

The Svalgaard modification of the sunspot number series would have to be widely accepted and further statistical tests would need to be done to establish the validity of the finding that the longer periods in the modified yearly sunspot data from 1700 to 2011 are all subharmonics of the main 11-year periodicity. It is worth noting though that subharmonics in the frequency spectrum would suggest a non-linear mechanism, possibly with chaotic instability [14].

4. References

[1] Lomb N R and Andersen A P 1980 Mon. Not. R. astr. Soc. 190, 723–32
[2] Waldmeier M 1961 The Sunspot Activity in the Years 1610-1960 (Zürich: Verlag Schulthess u. Co. AG)
[3] Lomb N R 1976 Astrophysics and Space Science 39 447–62
[4] Scargle J D 1982 Astrophys. J. 263 835–53
[5] Svalgaard L 2012 Comparative Magnetic Minima: characterizing quiet times in the Sun and stars: Proc. IAU Symp. 286 (Mendoza, Argentina, 3–7 October 2011) ed C H Mandrini and D Webb (Cambridge: Cambridge) 27–33
[6] Schwarzenberg-Czerny A 1991, Mon. Not. R. astr. Soc. 253 198-206
[7] Kolláth Z and Oláh K 2009 Astronomy & Astrophysics 501 695–702
[8] Waldmeier M 1935 Astron. Mitt. Zurich 14 105–30
[9] Gleissberg W 1939 The Observatory 62 158–9
[10] Dicke R H 1978 Nature 276 676–80
[11] Charbonneau P and Dikpati M 2000 Astrophys. J. 543 1027–43
[12] Charbonneau P 2010 Living Rev. Solar Phys. 7 3
[13] Basu S and Antia H M 2010 Astrophys. J. 717 488-95
[14] Adler L, Yost W T and Cantrell J H International Congress on Ultrasonics: AIP Conf. Proc. 1433 (Gdańsk, Poland, 5–8 September 2011) ed B B J Linde, J Pączkowski and N Ponikwicki (Melville: AIP) 527–30

Acknowledgments

An earlier poster paper version of this paper was co-authored with Martin Anderson of Sydney Observatory, whose assistance is gratefully acknowledged. Zoran Mikić of Predictive Science Inc suggested examining the sunspot data with the Svalgaard modification and helpful comments on the first draft of this paper were received from the referee, Robertus von Fay-Siebenburgen.