Search for a relationship between solar cycle amplitude and length

S. K. Solanki1, N. A. Krivova1,2, M. Schüssler1, and M. Fligge3

1 Max-Planck-Institut für Aeronomie, 37191 Katlenburg-Lindau, Germany
e-mail: natalie@linmpi.mpg.de; schuessler@linmpi.mpg.de
2 Astronomical Institute, St. Petersburg University, 198504 St. Petersburg, Russia
3 Institute of Astronomy, ETH Zentrum, 8092 Zürich, Switzerland
e-mail: fligge@astro.phys.ethz.ch

Received 13 August 2001 / Accepted 1 October 2002

Abstract. The cross-correlation between time series of solar cycle length and amplitude suggests that the length precedes the amplitude. The relationship between the two is found to be more complex than a simple lag or phase shift, however. A simple empirical model is constructed which allows the amplitude of a given cycle to be predicted with relatively high accuracy from the lengths of earlier cycles. This result not only adds to the means at our disposal for predicting the amplitudes of future cycles, but also implies that the solar dynamo carries a memory of the length of one cycle over into the next. It may also have a bearing on why solar cycle length correlates better with the Earth’s temperature record than cycle amplitude (Friis-Christensen & Lassen 1991). Thoughts on possible physical causes are presented.

Key words. Sun: magnetic fields – solar-terrestrial relations – sunspots

1. Introduction

The cyclic magnetic activity of the Sun exhibits a change in amplitude from one cycle to the next as well as a variation in cycle length. These two quantities are not quite independent, however, exhibiting an inverse correlation, as has been shown by, e.g., Dicke (1978) and Hoyng (1993). Hence stronger cycles tend to be shorter and vice versa, although with a large scatter.

This relationship represents an observational constraint on dynamo models of the generation of the Sun’s magnetic field, although not a very strong one owing to the scatter and the correspondingly small correlation coefficient of approximately −0.35. Encouragingly, models such as those of Hoyng (1993) and Charbonneau & Dikpati (2000) do indeed predict such a weak relationship. This relationship also has a bearing on the influence of solar variability on climate. One of the intriguing questions in this respect is why the correlation between the solar cycle length and northern hemisphere land temperatures (Friis-Christensen & Lassen 1991, 1994) is better than that between the latter quantity and cycle amplitude (Friis-Christensen & Lassen 1991). Thoughts on possible physical causes are presented.

2. Method, data and results

We analyse the Zürich relative sunspot number, \( R_z \), covering the period 1700 to 1999. The cycle amplitudes and lengths derived from the Zürich relative sunspot numbers, \( R_z \), in two ways are used. The first set of values is made available by the National Geophysical Data Center (NGDC, Boulder, USA) and is based on the work of Waldmeier (1961). For the cycle amplitude we use the official NGDC values, while the time interval between two activity minima is taken as the cycle length. The resulting amplitude and length curves of \( R_z \) are plotted in Figs. 1a and b, respectively, for cycles 1–22.

The other method of extracting cycle amplitudes and lengths from the \( R_z \) record involves applying the Morlet wavelet transform to the data set (Grossmann & Morlet 1984).
This technique has been applied to this and related data sets by Vigouroux & Delache (1993, 1994); Ochadlick et al. (1993) and Fligge et al. (1999) with the aim of determining the solar cycle length. The length and amplitude records of the solar cycle are derived from the position (frequency) and the amplitude of the main power ridge in the wavelet transform. These quantities can vary as a function of time. In Figs. 2a and b the amplitude and length records obtained by applying Morlet wavelets to $R_z$ are plotted. Note that the amplitude and cycle length values at the beginning and the end of a record are somewhat less accurate due to the property of the wavelet transform (cf. Fligge et al. 1999).

Figures 1c and 2c give the cross-correlation function between the corresponding amplitude and length records. Negative phase shifts imply that the cycle length record precedes the cycle amplitude.

Basically both figures (1c and 2c) exhibit a similar result. The most significant peak in the cross-correlation is negative, which corresponds to the well known result that shorter cycles tend to be stronger. The largest cross-correlation is not achieved at phase shift zero, however, but always at a negative phase shift. This corresponds to one cycle when the first technique is employed, which can only give discrete results, and to 11 years when wavelets are applied. The absolute value of the correlation coefficient then lies at 0.63 and 0.46, respectively.

We have calculated the probability that the largest deduced (absolute) correlations are due to chance, obtaining a probability
Fig. 3. Frames a) and b) show the same as Figs. 1c and 2c, respectively, but now based on the $R_z$ record restricted to the period after 1818.

below $10^{-3}$ for Fig. 1 and about $10^{-2}$ for Fig. 2. From these figures one could conclude that the cycle length precedes cycle amplitude by approximately one cycle, or that short cycles tend to precede strong cycles.

There is some doubt concerning the early part of the $R_z$ record due to data gaps and the non-uniform quality of the observations. Thus, prior to 1750 only yearly values of $R_z$ are available, between 1750 and 1818 only monthly values. Keeping in mind the improvement in the quality of $R_z$ data around 1750 and 1818 we have repeated the above analysis of $R_z$, but restricted to data obtained after 1750 and 1818, respectively. Restricting the data to the period after 1750 only leads to a slight increase in the anti-correlation, but if we consider just the higher quality data compiled after 1818, the first method of determining cycle length and amplitude gives the cross-correlation function plotted in Fig. 3a. The cross-correlation at a lag of $-1$ cycle now lies at the $-0.35$ level (the probability of a chance correlation is about 9%), while the strongest cross-correlation is achieved at a lag of $-3$ cycles. For the wavelet-based analysis, the minimum of the cross-correlation now reaches $-0.63$ at a lag of approximately $-22$ years (see Fig. 3b). Also, the width of the cross-correlation peak is significantly larger compared to the results from the longer time series. The results of applying the two techniques are roughly consistent, except that the Morlet wavelet, which introduces a smoothing into the determined cycle length record (cf. Fligge et al. 1999), also leads to a smoothed cross-correlation function.

From Figs. 1–3 one may either conclude that the relationship between the two records is simply a fluke or that it is more complex than Figs. 1 and 2 would suggest (and possibly is time dependent). To check for a time dependence we determined the cross-correlation between the length and amplitude of a series of partially overlapping 6-cycle long extracts from the $R_z$ record. We first take the length of the first 6 cycles (cycles 0–5) and cross-correlate this data set with the amplitudes of cycles 1–6 and of cycles 3–8. Then we shift the length record by a cycle and repeat the analysis, keeping the lags between the amplitude and length records fixed. In this manner we can roughly follow how the cross-correlation at lags of $-1$ and $-3$ cycles develops over time (Fig. 4). The cycle number in the abscissa refers to the middle of the extracted interval of amplitude. In the early record a lag of $-1$ provides by far the more significant cross-correlation, while at later times the lag of $-3$ became more significant. In both cases the cross-correlation is mainly negative, implying that shorter cycles are followed one or three cycles later by stronger cycles.

Again, two interpretations of this result are possible. Either we discount the earlier part of the $R_z$ record and consider only the result of Fig. 3 to be relevant, or we accept that the lag between the two records has been changing with time. Irrespective of which interpretation is preferred just shifting the cycle length record by 1 or 3 cycles does not give the best possible representations of the cycle amplitude record (see below). This is also the case for other shifts, as tests show. Note that $|r_{c}|$ for a lag of $-1$ cycle is largest for the period around the Dalton minimum, characterized by weak and long cycles. It may thus be that such a lag is typical only for periods of anomalously weak activity. If the proposal of Usoskin et al. (2001) is correct and cycle 4 is composed of 2 shorter cycles then the relation around this epoch would be completely altered. However, see Krivova et al. (2002), who have found considerable evidence against an additional cycle in the time frame 1790–1800.

Guided by Figs. 1–4 we model the cycle amplitude on the basis of the cycle length as described below. We then compare these modelled amplitudes with the cycle amplitude time
series derived directly from the measurements. The correlation between the two amplitude time series is a measure of the quality of the model. Let the cycle length record shifted by \( n \) cycles be denoted by \( l(n) \). We construct an approximation of the cycle amplitude record, \( a(0) \), as follows:

\[
a(0) = c[w(l(-n_1) (1 - w)) l(-n_2)] + d,
\]

where \( 0 \leq w \leq 1 \) is a weight, \( c \) is a factor and \( d \) an offset used to convert cycle length into amplitude. For \( n_1 \) and \( n_2 \) we tried various combinations of 0, 1, 2 and 3 and obtained best results for the combination \(-1\) and \(-3\), although the combination of 0 and \(-3\) cycles gave results that are only of slightly lower quality. A major disadvantage of the combination 0 and \(-3\) cycles is that it cannot be used to predict the amplitude of an upcoming cycle (see Sect. 4), so that in the following we concentrate on the combination \(-1\) and \(-3\).

After 1823 (cycle 7–22) we obtained the best result for \( w = 0.4 \) (weighting for a lag of \(-1\) cycle). For this period the correlation between the true and the modelled cycle amplitude record is 0.83. For the whole \( R_c \) record the correlation coefficient becomes 0.53 with this weight, but 0.71 with \( w = 0.7 \), which provides the best results for the whole time series.

A more detailed test of the model is provided by comparing shorter subsets, i.e. carrying out a similar analysis as underlies Fig. 4, except that we now compare the modelled and measured cycle amplitude. The resulting cross-correlations (at shift 0) are plotted in Fig. 5a. Specifically, we plot the correlation coefficient for subsets of 6 cycles of the cycle amplitude and the modelled cycle amplitude record. For the period prior to 1823, the best result is obtained for \( w = 1 \) (dot-dashed curve in Fig. 5, identical to the solid curve in Fig. 4), while for the period after 1823 the model with \( w = 0.4 \) (triple-dot-dashed curve in Fig. 5) shows a better agreement with the data \( (r_c = 0.83) \) than simply shifting the cycle length by 3 cycles (dashed curve in Fig. 4). If we attempt to obtain a good representation of the whole data set \( (w = 0.7 \); solid curve in Fig. 5), the correspondence, unsurprisingly is not as good at any given period, except around cycle 10, than for the 2 extreme cases discussed above. Constructing an artificial amplitude time series using a combination of three different lags did not lead to any significant improvement.

Yet another test is provided by employing the relationship deduced from cycles 22 and earlier to “predict” the amplitude of cycle 23. This value we can then compare with the official amplitude: 120.8. In Table 1 we list for the weights \( w = 0.7 \) and 0.4 the best estimates of the coefficients \( c \) and \( d \) (regression coefficients determined by searching for the highest correlation coefficient with the observed amplitudes), as well as the prediction for the amplitude of cycle 23, \( a(0)_{23} \), For both the weights the predicted amplitude lies within the estimated error margin of approximately 20.

Finally, led by the behaviour of the correlations in Figs. 4 and 5, we let \( w \) be time-dependent. For simplicity we assume that \( w \) varies sinusoidally between 1 and 0. There are two free parameters in this case, the period and the number of the cycle for which \( w = 1 \) is first reached. It turns out that there is no unique set of these parameters giving the best reproduction of the amplitude time series. In Fig. 6a we show an example of the weights for the \( l(-1) \) [solid; \( w(t) \)] and \( l(-3) \) [dashed; \( 1 - w(t) \)] records which give good results, in Fig. 6b the correlation coefficient for subsets of 6 cycles (similar to Fig. 5a). The correlation coefficient between artificial and real amplitude series is 0.82 for the whole interval, and 0.80 for the period since 1823. Note also that except for the sets of cycles centred on cycle 14 and on cycle 19 the correlation is always better than 0.7. Figure 6c shows the difference between the modelled and true cycle amplitude time series (compare with Fig. 5b). Of course, it is a matter of debate whether the improved representation of the whole time series using a time-dependent \( w \) is a worthwhile exercise. Possibly the older data are of insufficient quality and should not be considered. In this case the constant \( w = 0.4 \) model gives results of similar quality (Fig. 5).

![Fig. 5. a) Correlations between modelled and measured cycle amplitude (at fixed lag 0), as a function of time. Triple-dot-dashed curve: \( w = 0.4 \); solid curve: \( w = 0.7 \); dot-dashed curve: \( w = 1 \). The ordinate scale has been inverted in order to facilitate comparison with Fig. 4. b) Difference between actual and modelled cycle amplitude as a function of cycle number.](image)

<table>
<thead>
<tr>
<th>( w )</th>
<th>( c )</th>
<th>( d )</th>
<th>( a(0)_{23} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>-34.3</td>
<td>492.8</td>
<td>140.5</td>
</tr>
<tr>
<td>0.4</td>
<td>-27.7</td>
<td>419.9</td>
<td>119.6</td>
</tr>
<tr>
<td>( w(t) )</td>
<td>-30.6</td>
<td>453.9</td>
<td>120.6</td>
</tr>
</tbody>
</table>
Fig. 6. a) Time-dependent weight \( w \) given to the cycle length record shifted forward by one cycle (solid curve) and \((1 - w)\), the weight given to the record shifted by \(-3\) cycles (dashed). b) The same as Fig. 5a, but now for the time dependent \( w \) values plotted in frame a. Note the inverted ordinate scale. c) The same as Fig. 5b, but for time dependent \( w \).

We nevertheless present this analysis here, since it will become possible with the help of upcoming cycles to distinguish between both approaches: the predictions of both models will diverge from each other with time, as the relative weights change with time in one model, but not the other.

The last line of Table 1 shows how the model plotted in Fig. 6 fares when predicting the amplitude of cycle 23. It turns out that the model giving the best correlation with the amplitudes of cycles 0–22 \( r_c = 0.82 \) also provides the best (almost perfect) estimate for the amplitude of cycle 23. We believe that this is by chance, since other choices of \( w(t) \) which give \( r_c \geq 0.8 \) for cycles 0–22 provide estimates of \( a(0)_{23} \) that lie between 95 and 135. Again this range of \( a(0)_{23} \) is compatible with the errors obtained for earlier cycles.

Finally, one way of showing the improvement in predicting cycle amplitude is to plot the length-based prediction vs. the measured value (Fig. 7). In Fig. 7a, the length and amplitude of the same sunspot cycle are plotted versus each other, while in Fig. 7b the actually measured amplitude of the cycle is plotted versus the amplitude of that cycle predicted by the cycle length time series. Obviously the scatter of the points is significantly smaller in Fig. 7b than in 7a. For comparison, we plot in Fig. 7c the length of a cycle vs. the amplitude of the 3rd cycle after it. As expected, the scatter is not improved if we consider the whole period (all symbols; correlation \( r_c \approx -0.13 \), but is quite reasonable for data since 1823 (stars; \( r_c \approx -0.73 \); cf. Fig. 4).

The reason for considering a shift of \(-3\) cycles separately is that, given a significant relationship between \( a(0) \) and \( l(-3) \), it would be possible to predict (within a given uncertainty) the amplitude up to 2.5 cycles in advance. Most other indicators or techniques employed for predicting parameters of forthcoming cycles work 0.5 cycles or less ahead. We caution, however, that the accuracy of such a longer term prediction is lower than of a prediction employing Eq. (1). Nevertheless, the predictions for the amplitudes of cycles 23, 24 and 25 given assuming the relationship

\[
a(0) = c' + d'l(-3)
\]

are 96, 139 and 159, respectively. Hence, this relationship predicts a renewed increase in the general level of solar activity. The coefficients \( c' \) and \( d' \) are obtained from cycles 7–22.

3. Discussion

The relationship between cycle amplitude and the length of previous cycles possibly finds an interpretation in the framework of flux transport dynamo models of Babcock-Leighton type (see, e.g., Charbonneau & Dikpati 2000; Dikpati & Charbonneau 1999, and references therein). In this type of dynamo, the regeneration of the poloidal field is due to the emergence and subsequent transport of magnetic flux in active regions having a (rotationally induced) systematic tilt with respect to the east-west direction. The poloidal field is transported by the solar meridional circulation, first poleward near the surface and then inward to the radial shear layer (tachocline) at the bottom of the convection zone. There a strong toroidal field is generated, which is transported equatorward by the deep return flow of the meridional circulation.

The finite transport time of magnetic flux by the meridional circulation naturally introduces a time delay of the order of the circulation time. Dikpati & Charbonneau (1999) show that the cycle length is determined by the speed of the meridional flow: faster flow implies shorter cycles, and vice versa. Since the poloidal field decays in time owing to (turbulent) resistivity, a faster advective transport should also bring more poloidal flux to the tachocline, from which a larger amount of toroidal flux is generated. For instance, Makarov & Tlatov (2000) find a clear correlation between the dipole moment of the large-scale magnetic field during solar minimum with the amplitude of the following solar maximum (see also Legrand & Simon 1981, 1991; Layden et al. 1991; Sheeley 1991). In this way, we could possibly understand that strong cycles are typically preceded
by short cycles. It is unclear, however, how to fit a 3-cycle delay into this conceptual framework. This would require a memory of the system extending beyond two polarity reversals, for example a “conveyor belt” of three simultaneously migrating toroidal flux systems being carried one after the other by the equatorward flow at the bottom of the convection zone. At least regarding the magnetic polarities of the poloidal and toroidal flux systems, the conditions for a 3-cycle delay are the same as those of a 1-cycle delay, so that both would be in accordance with observation. However, it is difficult to envisage a strong change of the meridional flow velocity leading to a transition from one to three cycles in the conveyor belt without a concomitant change in the cycle period, which is not observed.

4. Conclusion

Cross-correlations between cycle amplitude and cycle length reveal that the solar cycle length record precedes solar cycle amplitude, although the exact relationship is complex and appears to change with time. Based on this lag information we have found a simple empirical relationship between cycle length and amplitude, so that using cycle length information of the preceding solar cycles it is possible to reproduce the amplitude of a given cycle with an average error of 20 in sunspot number.

This result may shed some light on the curious fact that the cycle length record correlates so well with air temperature above northern hemisphere land masses (Friis-Christensen & Lassen 1991), while cycle amplitudes reach a maximum after the (sea surface) temperature does (e.g., Fig. 1 of Reid 1987).

Our result gains in interest in connection with the recent findings of Solanki et al. (2000, 2002) that the secular variation of the Sun’s total and open magnetic field is affected by the solar cycle length. The total magnetic flux causes solar irradiance variations (e.g., Chapman et al. 1994; Solanki & Fligge 2002), while the Sun’s open flux is responsible for modulations of cosmic rays and has thereby been brought into connection with cloud cover and hence climate (Svensmark & Friis-Christensen 1997; Marsch & Svensmark 2000).

The relationship between cycle length and amplitude implies that the dynamo retains a very particular type of memory from one cycle to the next, since the following cycle remembers whether the previous or even earlier cycles were short or long. Evidence for such a memory has also been provided by the presence of a 22-year modulation of preferred longitudes of sunspot groups (Balthasar & Schüssler 1983; Balthasar & Schüssler 1984) and the cycle intensity (Mursula et al. 2001). Actually, the fact that the length of a given cycle affects the amplitude of the next and the third following cycle could be one factor in producing a 22-year periodicity. Models proposed by Hoyng (1993) and by Charbonneau & Dikpati (2000) do predict an anticorrelation between cycle length and amplitude (or between the phase shift and amplitude in the case of Hoyng’s analysis). Both sets of authors obtain the correct sign of the correlation and Charbonneau & Dikpati also reproduce the correct magnitude of the correlation coefficient. It remains to be tested to what extent these or other models also exhibit the more complex relationship between cycle amplitude and length shown by the observational data.

Using the cycle length may provide a novel way of predicting the amplitude of future cycles. A review of the different techniques employed to predict solar activity levels, in particular the sunspot number, has been given by Wilson (1994) and Conway (1998). The cycle length record is limited to predicting the amplitude of a cycle at the time of the preceding minimum. On the one hand this also gives a reasonable prediction of the sunspot number integrated or averaged over a cycle due to the good correlation between these quantities and cycle amplitude (Wilson 1988). On the other hand, sunspot minimum
can only be positively identified after the fact, so that the most reliable prediction is only possible approximately a year after the minimum.

We do not claim that using the cycle length record provides greater reliability than, e.g., precursor techniques (e.g., Thompson 1993) or statistical predictions using neural networks (see Conway 1998 and references therein). In particular, cycle length can only be employed to predict the amplitude of the next maximum and not its date or other parameters of the next cycle. Nevertheless, it is an independent indicator and when combined with other indicators and prediction techniques (Joselyn et al. 1997) may help to improve the accuracy of predictions. In one respect, however, the cycle length has an advantage over other techniques. As Fig. 7c shows, there is still a high correlation (correlation coefficient of $-0.73$) between the length of cycle $n-3$ and the amplitude of cycle $n$ (for data since 1823). Thus it is possible to predict the amplitudes 1.5 and 2.5 cycles ahead, with an accuracy of 25 in sunspot number.

Finally, note that the lags of 1 and 3 cycles are close to the time delay parameters of 9 and 32 years found by Orfila et al. (2002) using genetic algorithms. This provides independent support to our conclusions.

Acknowledgements. We thank an anonymous referee for critical comments on an earlier version of the paper, which greatly helped to improve it. We are also grateful to I. Usoskin for providing useful references.

References

Dicke, R. H. 1978, Nature, 276, 676
Friis-Christensen, E., & Lassen, K. 1991, Science, 254, 698
Wilson, P. R. 1994, Solar and stellar activity cycles (Cambridge Astrophysics Ser., Cambridge, MA: Cambridge University Press)