THE PHASE VARIATIONS OF THE SOLAR CYCLE

R. H. DICKE

Joseph Henry Laboratories, Physics Department, Princeton University, Princeton, NJ 08540, U.S.A.

(Received 10 July, in revised form 27 September, 1987)

Abstract. It has previously been shown that the statistics of the phase fluctuation of the sunspot cycle are compatible with the assumption that the solar magnetic field is generated deep in the Sun by a frequency stable oscillator and that the observed substantial phase fluctuation in the sunspot cycle is due to variation in the time required for the magnetic field to move to the solar surface (Dicke, 1978, 1979). It was shown that the observed phase shifts are strongly correlated with the amplitude of the solar cycle. It is shown here that of two empirical models for the transport of magnetic flux to the surface, the best fit to the data is obtained with a model for which the magnetic flux is carried to the surface by convection with the convection velocity proportional to a function of the solar cycle amplitude. The best fit of this model to the data is obtained for a 12-yr transit time. The period obtained for the solar cycle is $T = 22.219 \pm 0.032$ yr. It is shown that *the great solar anomaly* of 1760–1800 is most likely real and not due to poor data.

1. Introduction

The solar cycle is quite irregular with the period between sunspot maxima varying from 7 to 17 years. But statistically the distribution of periods favors the existence of a stable oscillator, with a variable but limited phase shift in the output signal (Dicke, 1978, 1979; Gough, 1980; Newkirk, 1984; Bracewell, 1985). The source of the solar magnetic field may be a high Q magnetofluid dynamic oscillator, deep in the Sun, and the variable phase shift may be due to variations in the time required for the generated magnetic flux to move to the solar surface.

The random walk in phase expected from a relaxation oscillator, an oscillator without a resonator, is not observed (Dicke, 1978). With such an oscillator, phase errors should be propagated indefinitely into the future.

A crude measure of the time required for the magnetic flux to move to the solar surface is obtained from the largest phase shift observed. This occurred about 1788 when the sunspot maximum arrived 6 years too early. If the above model is correct, substantially more than 6 years is required for the magnetic flux to move to the surface from the deep solar interior.

It has been shown that the phase shift in the output signal of the hypothetical solar cycle oscillator is not completely random but is strongly correlated with the variable amplitude of the 22 yr (Hale) sunspot cycle (Dicke, 1978, 1979). The sign of the correlation is consistent with the assumption that the magnetic flux moves to the surface most rapidly when the amplitude of the sunspot cycle is greatest. It was previously assumed that the phase shift is proportional to the amplitude of the solar cycle (Dicke, 1979).

In the absence of a working dynamic theory of the solar cycle, it is too much to expect a convincing quantitatively correct physical explanation of the observed statistical

Solar Physics **115** (1988) 171–181.

correlation. But general characteristics of acceptable transport mechanisms might be stated.

Transport mechanisms appear to fall in two broad classes: (1) where the flux is in isolated tubes which, by magnetic buoyancy, float upward through the fluid to the surface and (2) where the flux is carried upward in moving fluid, convected toward the surface.

Parameterized empirical models could be introduced to represent each of these general mechanisms and the parameters could be adjusted by least squares to give the best fit to the data. Also, these empirical models could be intercompared to see which gives the best fit to the data.

It must be emphasized that the existence of a satisfactory empirical model does not necessarily imply that an analogous physical model must exist. The first mechanism to be discussed is the 'magnetic buoyancy model'. The magnetic buoyancy of thin toroidal flux rings could cause them to float to the surface (Parker, 1955). Then the mean transit velocity would be expected to be a function of the magnetic pressure.

It might reasonably be assumed that the mean magnetic pressure is a monotonically increasing function of the amplitude of the solar oscillation at the time the flux is generated. This solar cycle amplitude is not observed at the solar surface until the flux arrives there. Thus the deviation of the flux transit time and the change in the solar cycle amplitude should occur synchronously. This could account for the correlation of phase shift with amplitude and it would give the correct sign for the correlation. But it is not obvious that there should be the previously assumed linear relationship between the phase shift and amplitude.

A more general assumption, that the mean transit velocity is proportional to a power of the amplitude of the solar cycle, A^{-s}, includes the linear relation as a special case. The transit time, or phase shift, should then vary as A^s. The solar cycle amplitude, A, should be evaluated at the time of arrival of the flux tube at the surface. (For a discussion of the calculation of A, see below.)

An alternative mechanism for transporting magnetic flux to the surface is provided by the 'convective model'. The flux might be transported upward convectively, such as by the large azimuthal convective rolls recently recognized in the solar velocity field (Snodgrass, 1987; LaBonte and Howard, 1980). It is assumed that the magnetic field strength of the flux carried convectively is relatively weak and that magnetic buoyancy becomes important only near the the solar surface, where the flux becomes concentrated in tubes and the field strength is large.

If the velocities of the convective rolls are affected by magnetic stresses at the solar surface the mean transit velocity might be a function of the solar activity, again measured by the amplitude of the solar cycle.

The dependence of the phase of the solar cycle on the amplitude of the cycle could provide useful information about the magnetic flux transport mechanism. Any satisfactory physical mechanism should give a relation between phase and amplitude compatible with the observations. Obviously this must also be true of a parameterized empirical model if it is to correspond to a believable physical model. But, as noted above, the fact that an empirical model works provides no guarantee that a satisfactory physical model of this type can be found.

In the discussion below the convection hypothesis and the buoyancy hypothesis are designated hypotheses H_1 and H_2, respectively.

2. The Sunspot Data

Eddy (1976) has emphasized that the annual mean sunspot numbers for the years prior to the discovery of the solar cycle may be questionable for they were constructed years later from the published literature and from archival maps. He has characterized the data for 1818–1847 as 'good' and for 1749–1817 and 1700–1748 as 'questionable' and 'poor', respectively.

One particularly worrisome point is that in part of this 'questionable' period the phase deviation is found to be remarkably large. Sonett (1983) has characterized this large irregularity of the solar cycles of 1760–1800 as *the great solar anomaly*. He has noted that in addition to the anomalously large phase shift during this period (Dicke, 1970) there is a large amplitude change.

The cause of the 18th century anomaly is not known. There are at least two possible explanations. The anomaly may be a real but not presently understood physical phenomenon or it may be due to poor data (Sonett, 1983). But whatever the cause, the anomalous data should be omitted before attempting to examine the relationship between the phase and the amplitude of the solar cycle.

If because of inadequate data the validity of the sunspot data of the late 18th century is questioned, then the data of the early 18th century must be especially questioned.

To avoid the questionable data the sunspot numbers prior to 1817 are ignored in making a least-square fit of a sunspot function to the data. The resulting fitted function can then be extrapolated backward to test the goodness of this fit to the data in the range 1700–1816.

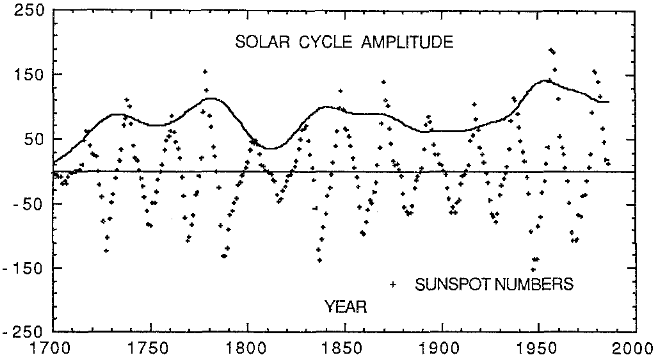

Fig. 1. The annual mean Hale sunspot numbers plotted as points and the computed amplitude of the solar cycle, $A(t)$, plotted as a curve.

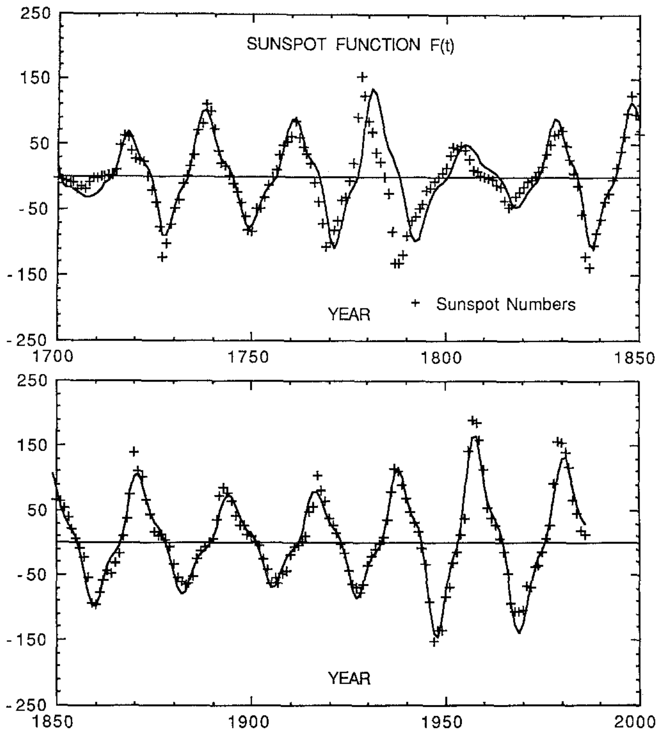

Fig. 2. The solar cycle function, $F(t)$, computed under hypothesis H_1 together with the Hale annual sunspot numbers.

To our great surprise it is found that the data of the first 6 decades of the 18th century are accounted for as well by this extrapolated fitted function as the 19th and 20th century data used to obtain the function. This suggests that the early data may be better than is sometimes thought and that the 'great anomaly' may be physically meaningful and not due to poor data.

The basic adopted data are the magnetic (Hale) annual sunspot numbers, $R(t)$, for which the signs of the sunspot numbers for successive (11 yr) half-cycles are changed (Bracewell, 1953). The years of sunspot minima yield the dividing points for sign change.

The procedure breaks down for the annual mean sunspot minima themselves. Here roughly equal contributions come from the new and old half-cycles and the separate contributions, which should be of opposite sign, are unknown. To avoid this difficulty the minimum sunspot numbers are redefined as the linearily interpolated values obtained from the preceeding year and the following year's numbers (of opposite sign).

For the years 1700–1960 the annual mean sunspot numbers are obtained from Waldmeir (1961) with the first 15 years corrected as suggested by Eddy (1976). For the years 1961–1986 a NOAA (1986) report is the source of the data and the 1986 value is an extrapolation. The sunspot numbers $R(t)$ are plotted as points in Figures 1 and 2. The 22 yr Hale sunspot cycle is evident.

3. The Fitted Sunspot Function

The form adopted for the calculated sunspot function, $F(t)$, includes both first and third harmonics of the frequency $\nu \cong (1/22)\,\mathrm{yr}^{-1}$. This provides a better fit to the data, $R(t)$, than the simple sinusoidal fit previously used (Dicke, 1979).

The solar cycle amplitude, $A(t)$, is first calculated from $R(t)$ as $a(t)$, a moving 23 year r.m.s. estimate of the amplitude:

$$a(t) = \sqrt{[(R^2(t-11) + R^2(t-10) + \cdots + R^2(t+11))/11.5]}\,. \tag{1}$$

$A(t)$ is obtained from $a(t)$ by filtering with a zero-phase shift low-pass filter, passing the frequency band $0 \leq \nu < 0.043\,\mathrm{yr}^{-1}$. Neglecting the small plus/minus asymmetry of the solar cycle, only odd harmonics of $\nu \cong 1/23\,\mathrm{yr}^{-1}$ would be expected in $R(t)$. But then the expected value of $A^2(t)$ is the sum of the squares of the amplitudes of these harmonics. This sum is predominantly the contribution from the first harmonic with only a small contribution from the third harmonic. The amplitude function $A(t)$ is plotted as a curve in Figure 1.

The computed sunspot function, $F(t)$, is given by the least-squares fit of $F(t)$ to $R(t)$, where $F(t)$ is given by:

$$F(t) = A(t)\{C_0 + C_1 \cos[2\pi\nu(t-1899) + \Phi(W(t,s,\tau) - \mathbf{W})] + \\ + S_1 \sin[\ldots] + C_3 \cos 3[\ldots] + S_3 \sin 3[\ldots]\}\,. \tag{2}$$

The phase term, $\Phi(W(t,s,\tau) - \mathbf{W})$, in the arguments of the trigonometric functions, is discussed below.

In the analysis it is assumed that in the period 1817–1986 the errors in $R(t)$ are normally distributed and that the variance, σ^2, is independent of t.

As noted above, the least squares fit of $F(t)$ to $R(t)$ is limited to the time period 1817–1986. To improve the convergence of the nonlinear fit of the frequency, ν, 1899 is chosen as the time zero in the arguments of the trigonometric functions.

Under hypothesis H_1, that the magnetic flux is conveyed to the surface in convective rolls, the convective motion might reasonably be assumed to be a function of the solar cycle amplitude, $A(t)$.

This function is assumed to be a constant plus a power, s, of $A(t)$. It will be assumed that the radial component of the fluid velocity is

$$\mathrm{d}r/\mathrm{d}t = f(r,\theta)\,[A(t)^s + C]\,. \tag{3}$$

Integrating (3) along a trajectory of the convective motion from the source of the magnetic field to the solar surface yields

$$\tau'[\{A(t)^s\} + C] = \int \mathrm{d}r/f(r,\theta(r))\,, \tag{4}$$

where τ' is the transit time and $\{A(t)^s\}$ is the mean of $A(t)^s$ over the transit interval, τ'. $A(t)$ is defined only for integral values of t and the above mean value is approximated as the average of the current value of $A(t)^s$ with the previous τ values. Here τ is the integer closest to the average of τ' over the period 1817–1986. In this period the phase fluctuation is observed to be of the order of 1–2 years, relatively small compared with the decade or more expected for the transit time (see above).

From Equation (4) the change in transit time, $\delta\tau$, due to 'switching off' the solar activity, i.e., setting $\{A^s\} = 0$, is

$$\delta\tau = \tau'\{A^s\}/C\,. \tag{5}$$

Assuming that $\delta\tau \ll \tau'$, $\delta\tau$ is proportional to $W = \{A^s\}$.

The phase shift term appearing in the brackets of Equation (2) is assumed to be $\Phi(W(t, s, \tau) - \mathbf{W})$, where the mean, $\mathbf{W}$, evaluated for the range 1817–1986 is included to improve the convergence in fitting for Φ. To find the optimum choice of the parameters Φ, s, and τ, they are adjusted by least squares along with the other parameters C_0, C_1, C_3, S_1, S_3, and ν. In the fit, τ is limited to positive integral values. The results for these least square fitted parameters are given in the first 2 lines of Table I, $H = 1$. The resulting function $F(t)$ is plotted as a curve in Figure 2 along with the sunspot numbers $R(t)$, plotted as points.

TABLE I

Least-squares fits[a]

H	C_0	C_1	C_3	S_1	S_3	ν (cycle yr^{-1})	Φ	s	τ (year)
1	0.051	0.051	−0.962	0.101	0.187	0.04501	0.0670	0.710	12
	±0.012	±0.019	±0.017	±0.018	±0.017	±0.00007	±0.0042		
2	0.056	0.062	−0.956	0.097	0.177	0.04521	−14.4	−0.350	(0)
	±0.014	±0.022	±0.019	±0.020	±0.020	±0.00007	±1.2		
3	0.059	−0.046	−0.942	0.128	0.096	0.04571	(0)		
	±0.018	±0.027	±0.026	±0.026	±0.026	±0.00008			

[a] See Equation (2).

The other hypothesis, H_2, that magnetic flux rings float buoyantly to the surface was discussed in the Introduction. The mean transit velocity, and reciprocal transit time, adopted there is proportional to A^{-s}, where A is to be evaluated at the time of arrival of the magnetic flux at the solar surface.

The above assumption, that under hypothesis H_2 the phase varies as A^s, does not require any change in the functional form of Equation (2), providing that the constraint $\tau = 0$ be imposed. It should be emphasized that the constraint $\tau = 0$ is imposed only to obtain the correct functional form for the phase term, $\Phi(W(t, s, \tau) - \mathbf{W})$ and does *not* imply that the average transit time is zero under hypothesis H_2. The parameters of the function F, fitted to $R(t)$ under H_2 are given in the third and fourth lines of Table I, $H = 2$.

Expressing the phase shift in time units, the transit time under H_2 is

$$\tau' = -\Phi A^s/2\pi\nu. \tag{6}$$

For the typical values $A = 50$, 100, and 150, $\tau' = 13.0$, 10.2, and 8.8, respectively.

It is of interest to compare the above two least-square fits with the fit made without the phase term present, hypothesis H_3. This least-squares fit is made with the constraint $\Phi = 0$. The results of this fit are given in the fifth and sixth lines of Table I, $H = 3$.

The errors given in Table I are only formal, treating $A(t)$ as an externally determined variable. Allowing for the degrees of freedom required to define $A(t)$, the above standard error estimates should be increased by $\sim 2\%$.

Under the hypothesis H_1 the term $\Phi(W(t, s, \tau) - \mathbf{W})$ seems adequately to describe the phase variation in the period 1817–1986. There is no indication of a random walk in phase, for the small phase deviations which appear are of short duration and are not propagated into the future. Also the large phase error found in the 'great solar anomaly' disappears abruptly.

4. Likelihood Ratios

We would like to have many sets of data which could be used to test an hypothesis. Instead we have only one set of solar data against which alternative hypotheses can be tested. The likelihood ratio of 2 alternative hypotheses is a convenient statistic which can be used to decide whether or not there is a statistically significant difference in the quality of the two fits.

The likelihood, L, of a fit of a function $F(t)$ to a data set $R(t)$ is conveniently defined through its logarithm

$$-2\ln(L) = N\ln(2\pi\sigma^2) + \chi^2. \tag{7}$$

Here $N = 170$ is the number of data points, and χ^2 and σ have their usual meanings. A maximum likelihood fit of $F(t)$ to $R(t)$ is obtained by varying the parameters C_0, C_1, C_3, S_1, S_3, ν, Φ, s, τ to minimize (7). Also σ is included as a parameter to be varied. The results obtained for the first 9 parameters are the same as those obtained from the least-square fit and the value obtained for σ is the r.m.s. residual of the fit. This implies that $\chi^2 = N$. The 3 values of σ obtained for the hypotheses H_1, H_2, and H_3, are, respectively, $\sigma = 14.324$, 16.313, and 21.964. The corresponding maximum likelihood value of L is given by the expression

$$L = (2\pi\sigma^2 e)^{-N/2}. \tag{8}$$

The ratio of likelihoods for two different hypotheses, e.g., H_1 and H_3, is

$$L_1/L_3 = (\sigma_1/\sigma_3)^{-N}. \tag{9}$$

From Equation (9) it is evident that the fits $H = 1$ and 2 are much better than $H = 3$ and that hypothesis H_1 is better than H_2.

An important question is whether or not there is a statistically significant difference between the fits under H_1 and H_2. This can be answered using Wilks theorem (Wilks, 1938). With the constraint $\tau = 0$, H_2 can be interpreted as the null hypothesis with H_1 as the alternative hypothesis.

The Wilks statistic is

$$W^* = -2\ln(L_2/L_1) = 2N\ln(\sigma_2/\sigma_1) = 44.21\,. \tag{10}$$

If the null hypothesis H_2 is true and the sample is large, W^* has a chi-squared distribution with 1 deg of freedom. A χ^2 as large as 44 is highly improbable and it is likely that H_2 can be rejected.

The residuals obtained under H_1 are found to be normally distributed. Using a chi-squared test, the distribution of residuals is divided into 17 blocks of equal expectation and a $\chi^2 = 12$ is obtained with 16 deg of freedom. Thus the hypothesis that the distribution is non-normal can be rejected.

5. The Great Solar Anomaly

The curve $F(t)$ plotted in Figure 2 shows that, under hypothesis H_1, the extrapolation of this fitted function backward 170 years to the first 6 decades of the 18th century yields a fit to $R(t)$ comparable with that obtained for the modern data. This is surprising in view of the characterization of these early data as 'poor'.

Subjective bias can probably be ruled out as the source of the agreement of $F(t)$ with $R(t)$ for these early data. The frequency, ν, of the underlying solar oscillator and the phase amplitude relation were unknown when the data were compiled. As an aid in comparing the early and late data the squares of the residuals, $(R(t) - F(t))^2$, are plotted as a curve in Figure 3. To show the increase in variance under hypotheses H_2 and H_3, the squared residuals for the second and third fits of Table I are plotted in Figures 4 and 5.

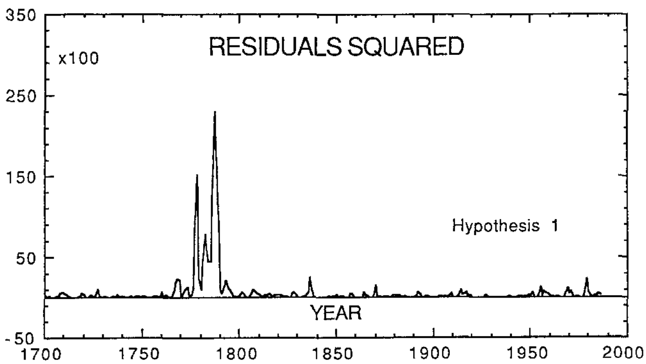

Fig. 3. The square of the residuals obtained from Figure 2. The 'great solar anomaly' toward the end of the 18th century is obvious. Surprisingly the fit prior to the anomaly is as good as that for the modern data.

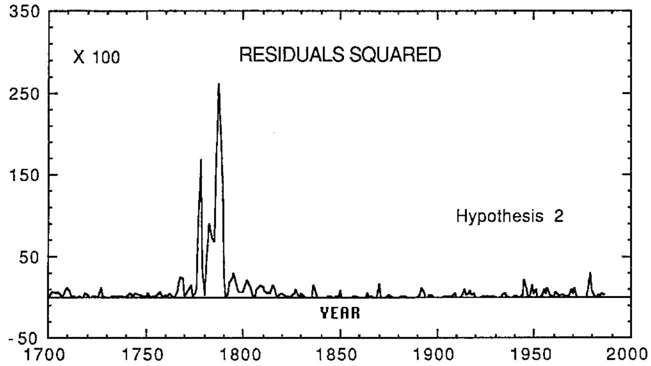

Fig. 4. The same as Figure 3 but computed under H_2.

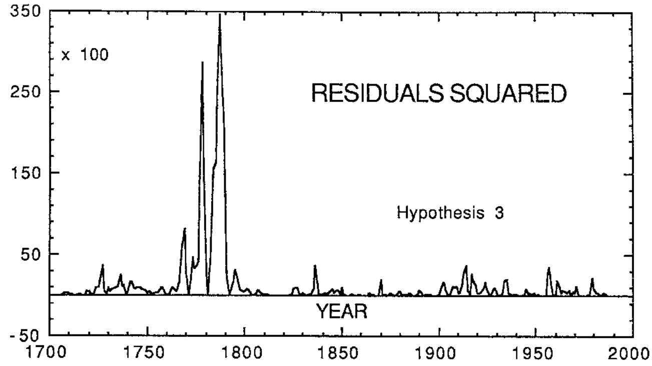

Fig. 5. The same as Figure 3 but computed under H_3. Under H_3 the amplitude correlated phase correction is omitted. Comparing Figures 3 and 4 with Figure 5 shows the large increase in the residual variance obtained when the phase correction is omitted from $F(t)$.

It should be noted that, for all 3 sets of residuals the quality of the fits for 1700–1760 is approximately the same as for 1817–1986. This implies that these early data have essentially the same phase-amplitude relation as that found for the modern data.

The large phase anomaly of the sunspot cycle during the last 4 decades of the 18th century, and its quick disappearance at the beginning of the 19th century have been discussed above. If this is the result of using 'questionable' data one might expect the larger phase errors to be found in the period 1798–1822 when the solar cycle was anomalously weak and errors relatively more important. Instead the largest phase errors occurred when the amplitude was anomalously large. This tends to support the suggestion that the phase anomaly is real and not due to poor data.

The values of the mean square residuals for H_1, H_2, and H_3 in the interval 1700–1760 are respectively 158, 253, and 599. For the interval 1760–1800 the corresponding values are 3390, 4340, and 7010. For the interval 1800–1986 the values are respectively 214, 338, and 461. Note that the second set of variances is an order of magnitude greater than the third, but that the 3 ratios of the 2 sets are respectively 15.9, 12.8, and 15.2.

The approximate constancy of this ratio shows two things: (1) that the relation between phase shift and amplitude seen in the 19th and 20th centuries is also present during the solar anomaly and (2) that the variance of the residuals is proportional to the variance of the signal due to phase error. This latter could be true if the noise component of the residual variance was negligible in comparison with an unmodeled part of the signal and this could be true if the phase error was imperfectly modeled in Equation (2). Much of the variance of the residuals may be unmodeled signal, rather than noise.

6. Summary and Conclusions

It is clear from Figure 3 that the fit of the function $F(t)$ to the data in the first 6 decades of the 18th century is as good as the fit for the period 1817–1986 used to obtain the function. This is evidence that the large phase anomaly occurring in the interval 1765–1800 was not primarily due to poor data but rather represents some presently unknown physical phenomenon.

The hypothesis H_1, that the mean transit velocity of magnetic flux to the solar surface is proportional to the sum of a constant and a power, s, of the amplitude of the solar cycle averaged over the transit interval yields the best fit to the sunspot numbers for an average transit interval (integer) of $\tau = 12$ years and an amplitude power of $s = 0.71$. The resulting fit to the data is substantially better than the fit under hypothesis H_2, for which the transit velocity is proportional to a power, $-s$, of the amplitude evaluated at the time of arrival of the flux at the surface. The best fit under H_2 is obtained for the power $s = -0.35$. This fit implies that under H_2 the transit time is about a decade, 10.2 years for $A = 100$.

In the original analysis (Dicke, 1979), it was assumed that the phase shift was proportional to the amplitude, i.e., $s = 1$. This yields $\sigma = 16.784$. This is only slightly larger than the value $\sigma = 16.313$ obtained under H_2.

It is evident from Figures 2 and 3 that, except for *the great solar anomaly*, the fit to the sunspot data is remarkably good under hypothesis H_1 and this fit yields $T = 22.219 \pm 0.032$ years as the period of the solar cycle. It must be emphasized, however, that this satisfactory fit does not necessarily imply the correctness of the underlying physical model which originally suggested the hypothesis H_1. Under this model the magnetic flux is transported to the surface convectively, but the assumption that the convective velocity is proportional to a constant plus a power of the solar cycle amplitude averaged over the past τ years is quite arbitrary. Also there could be other mechanisms for which the phase shift is proportional to a power of the amplitude averaged over the transit interval in the past.

Acknowledgement

This research was supported in part by the National Science Foundation.

References

Bracewell, R. N.: 1953, *Nature* **171**, 649.
Bracewell, R. N.: 1985, *Australian J. Phys.* **38**, 1009.
Dicke, R. H.: 1970, in A. Slettebak (ed.), *Stellar Rotation*, D. Reidel Publ. Co., Dordrecht, Holland.
Dicke, R. H.: 1978, *Nature* **276**, 676.
Dicke, R. H.: 1979, *Nature* **280**, 24.
Eddy, J. A.: 1976, *Science* **192**, 1189.
Gough, D.: 1980, in S. Sofia (ed.), *Variations of the Solar Constant*, NASA Conf. Report 2191.
LaBonte, B. and Howard, R.: 1980, *Astrophys. J.* **239**, L33.
Newkirk, Jr., G.: 1984, in *Proc. 3rd Symposium on Accelerator Mass Spectroscopy*, Zurich.
NOAA: 1986, *Solar Geophysical Data*, No. 505, part I, NOAA, Boulder, Colorado, U.S.A.
Parker, E. N.: 1955, *Astrophys. J.* **121**, 491.
Snodgrass, H. B.: 1987, *Astrophys. J.* **316**, L91.
Sonett, C. P.: 1983, *Geophys. Res. Space* **88** (A4), 3225.
Waldmeir, M.: 1961, *The Sunspot Activity in the Years 1610–1960*, Schulthess, Zurich.
Wilks, S.: 1938, *Ann. Math. Stat.* **9**, 60.