Aust. J. Phys., 1985, **38**, 1009–25

Sunspot Number Series Envelope and Phase*

R. N. Bracewell

Electrical Engineering Department,
Stanford University, Stanford, CA 94305, U.S.A.

Abstract

The sunspot number series $R(t)$ from 1700 to date is found to be representable by $R(t) = |\mathscr{H}\{\mathrm{Re}(E(t)\exp[\mathrm{i}\{\omega_0 t + \phi(t)\}]) + U(t)\}|$, where ω_0 is the angular frequency corresponding to a period of 22 years, $E(t)$ is the instantaneous envelope amplitude, $\phi(t)$ is the instantaneous phase of a complex time-varying analytic function, $U(t)$ is an undulation of low amplitude and period about 30 (22-year) cycles and $\mathscr{H}$ is a nonlinear operator whose main effect is to introduce a small amount of third harmonic (period about 7 years). The justification for the 22-year period is the known fact that the observable sunspot magnetic fields reverse polarity every 11 years or so at the time of sunspot minimum; the undulation has been demonstrated, and its period determined, in fossil records discovered by Williams; and the third harmonic is an expected consequence of minor nonlinearity in the dependence of the arbitrarily defined $R(t)$ on the physical cause of sunspots. The algebraic representation is established by the Hilbert transform method of forming a complex analytic function as proposed by Gabor. The method reveals three obscuring features that may be alleviated as follows: use the alternating series $R_{\pm}(t)$ in which alternate 11-year cycles take opposite signs, remove the third harmonic, and subtract the undulation. These justifiable steps remove artificial components, such as sum and difference frequencies, that are gratuitously and nonlinearly introduced by conventional Fourier analysis as applied to the rectified, or absolute, value of the 22-year oscillation. Then a complex envelope $E(t)\exp\{\mathrm{i}\,\phi(t)\}$ can be discerned whose intrinsic behaviour can be studied to reveal statistics that bear on the physical origin of the solar cycle. The results favour a deep monochromatic oscillator whose influence is propagated to the observable surface via a time-varying medium. The r.m.s. value of the component of $E(t)$ is 0·4 of the mean and the characteristic time is a century. Frequency analysis of the envelope does not support a 78-year period in the modulation noticed by Wolf. Both the statistical frequency distribution of the amplitude $E(t)$ and its spectrum are subject to refinement by analysis of fossil solar records. The results do not favour the theory that the 22-year period is set by the natural frequency of a resonator with characteristic damping subject to random turbulent excitation. Also disfavoured is the theory of energy release at intervals determined by a relaxation process. Correlation has been found between the phase departure $\Delta\phi(t)$ from linear and envelope amplitude and attributed to propagation of the magnetic cycles through a time-varying, such as a convecting, medium. A correlation not depending on Hilbert transform analysis is predicted between the reciprocal cycle length and envelope amplitude and found to exist. Variability of the sunspot cycle length can be viewed as a Doppler shift due to propagation in a time-varying medium and the Wolf modulation then represents the concomitant intensity change. Agreement has been found between $E(t)$ and $\phi'(t)$ but not explained. If the explanation is dispersion in the propagation of the assumed magnetic flux waves then there is a mode of oscillation that has the characteristics required for the undulation $U(t)$. Extra buoyancy possessed by the magnetic field of strong cycles accounts for the fast rise time of strong cycles.

* Paper presented at the R. G. Giovanelli Commemorative Colloquium, Part II, Tucson, Arizona, 17–18 January 1985.

0004-9506/85/061009$02.00

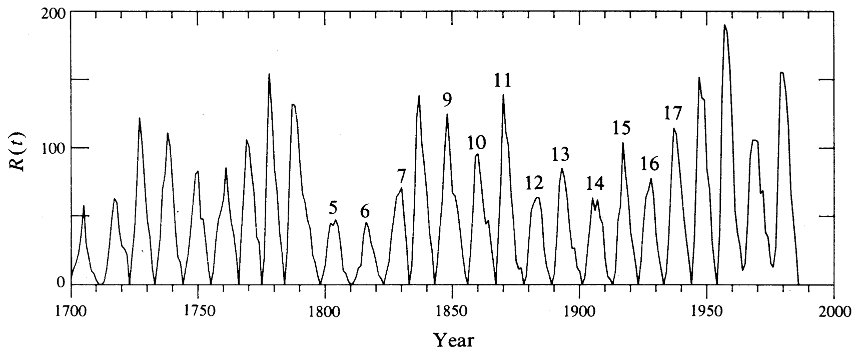

Fig. 1. Sunspot number series from 1700.

1. Introduction

The solar cycle has a period of about 22 years made up of two 11-year sunspot cycles whose properties have been summarized by Giovanelli (1984). Maxima of the sunspot cycle vary by a factor of more than four to one in sunspot number and the locus of maxima exhibits structure that has been commented on from the time of Wolf (1862). One conspicuous feature is the alternation seen from cycle 9 to cycle 17 in Fig. 1. Another is the persistence of low maxima for three consecutive cycles (5 to 7) and of high maxima at other times. This amplitude modulation, which I shall refer to as the Wolf modulation, has been described as exhibiting a period of about 80 years. The reports of fossil solar cycles by Williams (1981, 1983) also bear on the modulation and, because the number of years of records greatly exceeds what is available from modern observations, the fossil data will be of great importance.

The phase of the sunspot oscillator has been less discussed. An example is the paper by Dicke (1978) in which it is asked whether there is a chronometer deep in the Sun rather than a relaxation oscillator.

Study of the envelope of sunspot maxima is hindered by the fact that maxima are accessible only at 11-year intervals. Phase is determinate only at turning points; what happens in between times to the envelope and to phase is not clear. Of course the samples that become available every 11 years or so are clearly affected by noise, partly observational and partly solar; it would be desirable to incorporate all the observed sunspot numbers into a determination of amplitude and phase so that random noise could be reduced by averaging.

This paper presents a method of analysis of annual time series containing solar cycles. It is applicable both to sunspot number counts and to varve thickness measurements.

2. Envelope and Phase

There is a well-known mathematical technique for the extraction of instantaneous envelope amplitude and instantaneous phase from an oscillating quantity (Bracewell 1978) by use of the Hilbert transform, and apart from applications in electrical communications and information theory (Gabor 1946) the method is familiar in optics (Born and Wolf 1959). The idea is to represent a real oscillating quantity by a complex one whose real part varies with time as given. A special case is the use of a complex phasor to represent alternating current. Let $R(t) = \langle R\rangle + R_{\text{fluc}}(t)$ be the sunspot number, taken as a real function of continuous time. Then we associate with the fluctuating part $R_{\text{fluc}}(t)$ of $R(t)$ an imaginary time function $I(t)$ which is the Hilbert transform of $R_{\text{fluc}}(t)$:

$$I(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{R_{\text{fluc}}(t')\,\mathrm{d}t'}{t'-t}.$$

Since the Hilbert transforms of $R(t)$ and $R_{\text{fluc}}(t)$ are the same the subscript may be dropped. From the Hilbert transform we generate the complex function

$$R(t) - \mathrm{i}I(t).$$

Then the instantaneous envelope amplitude $E(t)$ is given by

$$E(t) = [\{R(t)\}^2 + \{I(t)\}^2]^{\frac{1}{2}},$$

and the instantaneous phase is given by

$$\tan\phi = -I(t)/R(t).$$

The original real waveform is expressible in terms of the complex representation $E(t)\exp\{\mathrm{i}\,\phi(t)\}$ as

$$R(t) = \langle R\rangle + \mathrm{Re}\{E(t)\,\mathrm{e}^{\mathrm{i}\phi(t)}\,\mathrm{e}^{\mathrm{i}\omega_{11}t}\},$$

where ω_{11} is a reference frequency that would be chosen as the angular frequency corresponding to the period of 11 years. The Hilbert transform formalism has become familiar in optics as an effective way of handling quasi-monochromatic waveforms. Envelope and phase statistics, coherence functions and other topics may be handled in this way. Although it is a far cry from the terahertz frequencies of optics to the nanohertz range of the sunspot oscillation, there is no apparent reason why the sunspot cycle should not be subjected to Hilbert transformation to establish the associated imaginary quadrature function $I(t)$.

3. Hilbert Transform of Sunspot Series

The standard sunspot series $R(t)$ of Fig. 1 was subjected to numerical Hilbert transformation and it was immediately found that the result does not lead to a satisfactory interpretation in terms of an envelope. One reason is that there is both an upper and a lower envelope to an oscillation and the sunspot series has the peculiar feature that the lower envelope is more or less flat, not partaking of the modulation of the upper envelope. In addition, the maxima and minima are not equally rounded. These two sorts of asymmetry do not arise in fields where envelope statistics have been successfully studied by the use of the Hilbert transform and may perhaps block any attempt at envelope analysis by means of the complex representation, even though the complex function is readily determined.

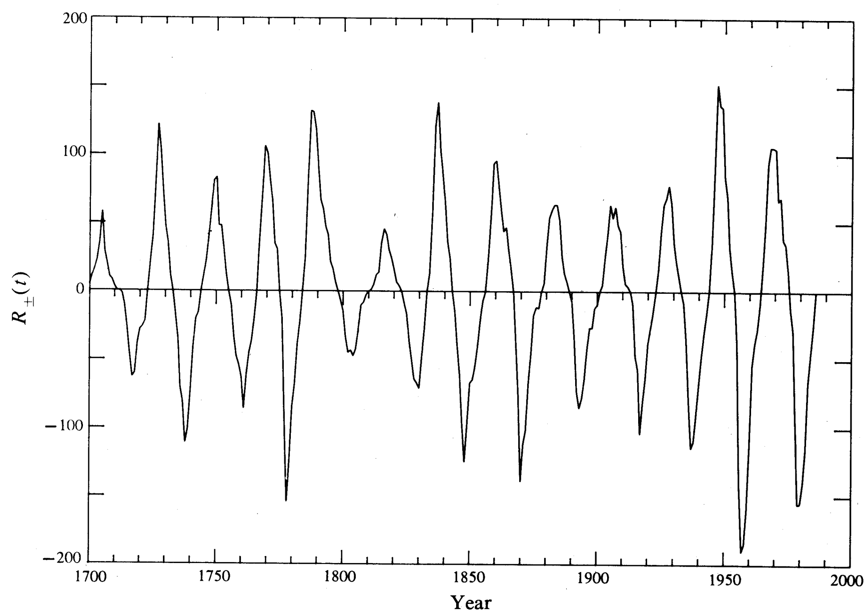

Fig. 2. Sunspot series presented as a 22-year oscillation $R_{\pm}(t)$ with symmetry about the mean.

4. Alternating Representation

There is however an alternative way of thinking about the sunspot series in which alternating signs are given to successive 11-year cycles (Bracewell 1953), as shown in Fig. 2. The reasoning behind this modification is absolutely fundamental; as discovered by Hale in 1913 (Giovanelli 1984) the magnetic period is 22 years. Of course, the sunspot number rarely drops to zero at sunspot minimum whereas the alternating cycle always passes through zero. However, if spots in the next cycle are to be given the opposite sign then there is a sense in which the sunspot number does pass through zero at the moment when spots of the old cycle sink through the level of equality with the oppositely signed spots of the new cycle that is rising. In Fig. 2 the minima were replaced by zero for this reason. Various authors have found the alternating cycle convenient (Hartmann 1971; Sonett 1984), but the idea has not been widely adopted.

In the present application, however, the centre-zero symmetrical oscillation is a most helpful adaptation to the method of envelope analysis now proposed. The notation $R_{\pm}(t)$ is used to distinguish the alternating series from the standard series $R(t)$.

When the Hilbert transform was again computed, this time for $R_{\pm}(t)$, two further features assumed significance. The first of these is the alternation previously referred to in cycles 9 to 17, where successive maxima alternate between high and low. The distinctive epithet 'zig–zag' was used by Williams (1981) to describe this alternation. The second feature is dealt with in Section 6.

5. Zig–Zag Effect

The zig–zag effect, if it could be removed, would facilitate envelope analysis; if left in, the effect requires the envelope to act with extra agility that puts a strain on

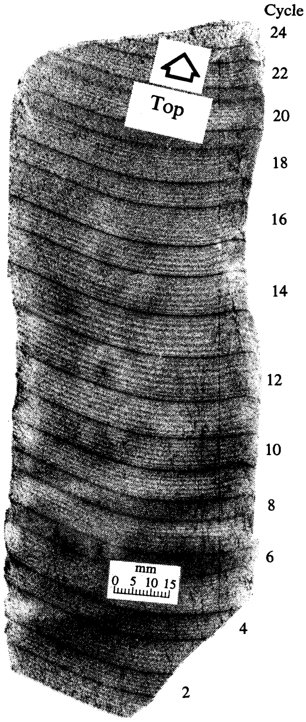

Fig. 3. A varved rock from the Elatina formation at Pichi Richi Pass, South Australia, exhibiting an effect connected with the sunspot zig-zag phenomenon.

Fig. 4. An artificial varve pattern generated for the years 1907 to 1927. The square waveform shown has one cycle per year with duration proportional to the mean sunspot number for that year. Ink is deposited seasonally to produce annual layers of thickness proportional to the year's sunspot number.

the computation and impedes averaging. In the description below the zig–zag effect may be put back in, but by addition rather than as a feature of the envelope.

A clue to how to remove the zig–zag effect is provided by Fig. 3. The fine bands are annual deposits on a periglacial lake bed (Williams 1981) which vary in thickness in cycles of 9 to 14 years. Where the bands are thin there is a conspicuous dark zone due to the close packing. This piece of rock was in my possession for some time on loan from Giovanelli, with whom I discussed the question whether the dark zones should be associated with sunspot maximum or minimum. Assuming one accepts the solar connection, there is still a question whether the annual layer thickness is correlated with or anticorrelated with sunspot number. This question is resolved by Fig. 4 in which an artificially varved pattern is generated from a solar cycle on the basis that the deposited layers are simply proportional in thickness to the sunspot number for the year. The resulting pattern acceptably mimics the real rock in a way that is tolerant to changes in cycle length or strength. An inverse dependence of layer thickness on sunspot number cannot be found that will explain the rock appearance; the reason is the phenomenon referred to in Section 3, namely that the upper envelope is variable while the lower envelope is practically flat. Using the dark zones as cycle separators, we can number the cycles sequentially as in Fig. 3. Cycles 2, 4, 6 and 8 are distinctly narrower than the intermediate cycles. This corresponds to a sunspot zig–zag where the even-numbered cycles are weaker than the odd-numbered ones. In a strong sunspot cycle the 11 or so layers deposited would be thicker and so the width between dark zones would be greater, as is seen with cycles 1, 3, 5 and 7. Now, when we look further up in the rock layers the inequality dies out; for example, cycles 12, 13 and 14 are of almost equal thickness, but higher still the effect returns, being most marked near cycle 21.

When the alternation returns it does so in antiphase. Whereas even cycles were thin near cycle 6 it is the odd ones that are thin near cycle 21. The interpretation of this in terms of the alternating cycle representation of Fig. 2 is that an additive undulation is present of half period about 15 cycles and amplitude (expressed in the sunspot number scale) of about 10. A physical interpretation of the zig–zag effect has been offered by Sonett (1984) in terms of a relict magnetic field; the relict would have to be alternating to be compatible with Fig. 3. The period T_U is not connected with any other known solar phenomenon.

Understanding that an undulation is present, as clearly revealed in the Elatina formation (Fig. 3), and stipulating that the sunspot cycle is behaving in the same way now as in the remote past, one can modify Fig. 2 by subtracting a term

$$U(t) = -10\cos\{2\pi(t-1910)/T_U\},$$

where T_U, the period of the undulation, is about 30 (11-year) cycles. The effect is to neutralize the zig–zag effect where it is pronounced. Some effect is also introduced near the beginning of the series, where other variations in the envelope are also present, and the impact is less obvious.

6. Third Harmonic

Fourier analysis of the standard sunspot number series $R(t)$ reveals numerous harmonics of the 11-year fundamental and many other non-negligible coefficients. The spectrum is not easy to interpret, partly because the series is not very long, contains a substantial random component, and is subject to uncertainty in the earlier years. In addition, there is much spectral complication due to the 'full-wave rectification' that distinguishes the standard series $R(t)$ from the alternating representation. This rectification introduces sharpish corners at the minima that require higher harmonics, and the forcing of the lower extrema onto a flattened line introduces other spectral detail. When the alternating cycle $R_{\pm}(t)$ is Fourier analysed the results are much simpler. The 11-year cycle practically disappears in favour of the 22-year fundamental and all the detail associated with rectification drops out. What remains is intrinsic to the solar cycle (and its noise) and free from artefacts associated with the mode of presentation. The most conspicuous spectral item apart from the fundamental is the peak at the third harmonic.

When the Hilbert transform of a waveform containing the third harmonic is taken, the resulting quadrature function is not suitable for forming a complex function from which an envelope can be extracted that satisfies intuitive expectations. The envelope generated is certainly tangential to both the data curve and the transform, but it also contains regular oscillations that have nothing to do with the slower changes in amplitude that we wish to bring out, and it would be better if such oscillations were suppressed. The envelope oscillations can be smoothed out *ad hoc*, but a more fundamental approach is to consider what causes the third harmonic. Sunspot number $R(t)$ is defined as $k(10g+f)$, where g is the number of sunspot groups on the visible hemisphere, f is the number of spots, and k is a personal factor characterizing the observer (around 0·6 in recent times). Both g and f have subjective components that depend on seeing conditions but, averaged over time, sunspot number correlates well with total sunspot area for the hemisphere and has generally been accepted as a convenient substitute for more fundamental physical measures such as total area. Nevertheless, long-lived mature sunspot groups that contribute about 8 ± 1 to the sunspot number differ in their contribution to physical area by more than an order of magnitude, for example from 10^{-5} to 3×10^{-4} of the hemisphere. Consequently, there is no reason to think that sunspot number as defined should depend linearly on any causal physical variable. Departure from linearity will, when the cause of the sunspots is a sinusoidally varying physical quantity, result in the sunspot number exhibiting an in-phase third harmonic. The logical approach is to recognize the likely origin and remove the third harmonic at the start. This harmonic is not strong, and is not an important constituent of the envelope modulation, although its relative strength may be found to be connected with the modulation envelope when this envelope has been determined.

To remove the third harmonic it would suffice to shift the alternating series $R_{\pm}(t)$ one-sixth of a 22-year cycle and add it to itself. This would cancel the harmonic and introduce a little smoothing. The method actually used was to combine twice $R_{\pm}(t)$ with the sum of the values occurring 3 and 4 years later by the expression

$$\tfrac{1}{4}\{2R_{\pm}(t)+R_{\pm}(t+3)+R_{\pm}(t+4)\}\,.$$

The transfer function $T(f)$ for this operation is $\frac{1}{4}\{2+\exp(\mathrm{i}\,2\pi 3f)+\exp(\mathrm{i}\,2\pi 4f)\}$. At the third harmonic, we have $f = 3/22$ cycles per year and $|\,T(3/22)/T(0)\,| = 0{\cdot}136$, so the third harmonic is attenuated by a factor 0·136. At the fundamental, where $f = 1/22$, there is also some attenuation, by a factor of $|\,T(1/22)/T(0)\,| = 0{\cdot}873$. While this procedure is not perfect, it reduces the third harmonic to a level that makes the Hilbert transform analysis acceptable. The finite-sum formula for attenuating the third harmonic introduces a time delay of 1·75 yr, which may be corrected by linear interpolation. The corrected coefficients are calculated by the convolution

$$\tfrac{1}{4}\{2\,0\,0\,1\,1\} * \{\tfrac{3}{4}\,\tfrac{1}{4}\} = \tfrac{1}{16}\{6\,2\,0\,3\,4\,1\}\,.$$

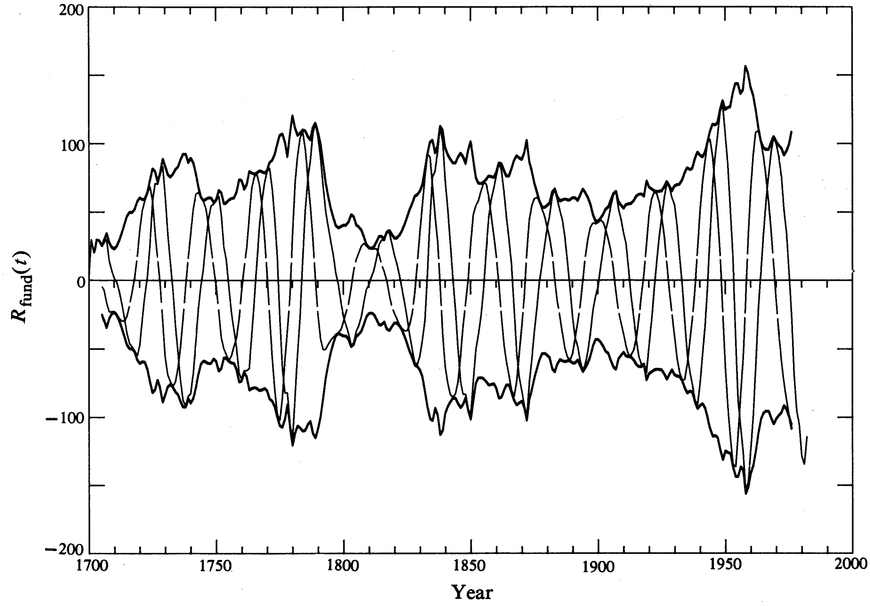

Fig. 5. Modified sunspot series $R_{\rm fund}(t)$, its Hilbert transform (in quadrature, lagging) and the corresponding top and bottom envelopes (heavy curves).

We are led to a modification of the alternating cycle that could be described as the undulation-free, harmonic-depleted fundamental $R_{\rm fund}(t)$, given by

$$R_{\rm fund}(t) = \frac{6R_{\pm}(t-2)+2R_{\pm}(t-1)+3R_{\pm}(t+1)+4R_{\pm}(t+2)+R_{\pm}(t+3)}{16\times 0{\cdot}873} - U(t)\,.$$

The divisor $0{\cdot}873$ is included to make $R_{\rm fund}(t)$ comparable with $R_{\pm}(t)$. A graph of $R_{\rm fund}(t)$ is shown in Fig. 5 together with its Hilbert transform $I_{\rm fund}(t)$ and the corresponding envelope $E(t) = (R^2_{\rm fund} + I^2_{\rm fund})^{\frac{1}{2}}$. The resulting envelope is relatively free from artificial 11- or 22-year hum and is sufficiently noise-free to be interpretable without further smoothing.

7. Instantaneous Phase

From the Hilbert transform one also obtains the instantaneous phase $\phi(t)$ by forming the complex analytic representation

$$R_{\rm fund}(t) - {\rm i}\, I_{\rm fund}(t) = E(t)\,{\rm e}^{{\rm i}\phi(t)}\,.$$

Then we get

$$\phi(t) = \arctan\{-I_{\rm fund}(t)/R_{\rm fund}(t)\}\,.$$

A graph of $\phi(t)$ is shown in Fig. 6. It is 'clean' enough to interpret without further treatment and may be compared with the straight line representing a linear phase advance of exactly thirteen 22-year cycles over the period 1700 to 1988. To facilitate study of the phase variations we also show, on an expanded scale, the departure $\Delta\phi(t)$

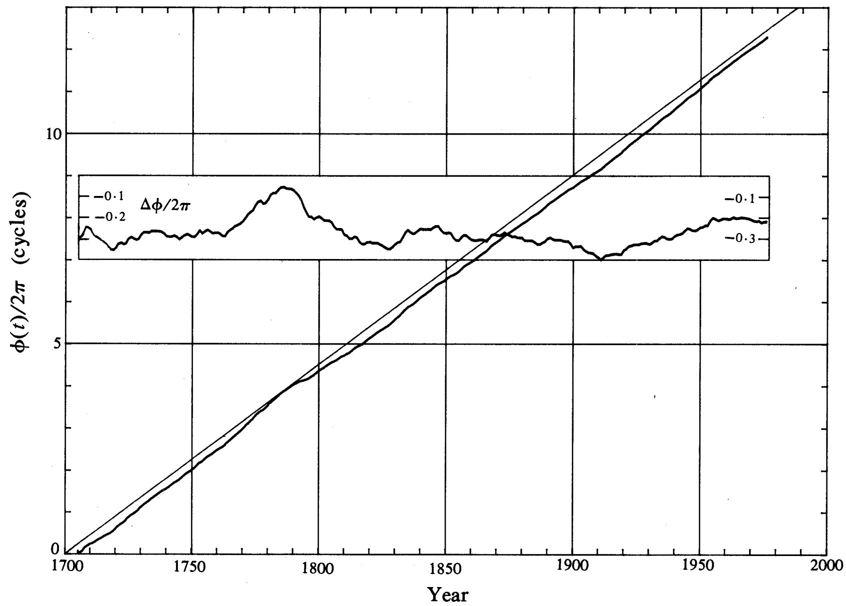

Fig. 6. Instantaneous phase $\phi(t)$ and, on an expanded scale, the departure $\Delta\phi(t)$ from the straight line.

from the arbitrary reference straight line representing a phase advance beginning at zero in 1700 and progressing regularly through 13 turns by 1988. Thus we have

$$\Delta\phi(t) = \phi(t) - \phi_{\text{ref}}(t),$$

where

$$\phi_{\text{ref}}(t) = 26\pi \frac{t-1700}{1988-1700}.$$

8. Complex Representation

In Fig. 7 the complex sunspot series, after removal of the undulation producing the zig–zag and the third harmonic, is shown on the complex plane with time as a parameter. As may be seen, the locus has a more or less circular form that expands and contracts regularly and is pricked out at a rather regular angular rate. There are just over twelve turns.

If the Elatina undulation were included, the whole pattern would drift slowly left and right by a small amount through about one oscillation during the whole time period. Clearly the composite description has simplified the study of the total locus. If the third harmonic were included, the circular shapes would become slightly elliptical with inclined axes. The radius vector would then oscillate between minor and major values with an 11-year period, a phenomenon not reflecting the slower size changes associated with amplitude modulation but connected rather with the waveshape of the 11-year cycle.

In the foregoing discussion the sunspot number series has been broken down into constituent parts. The operation needed to reconstitute $R_{\pm}(t)$ from its parts is

$$R_{\pm}(t) = \mathscr{H}\left\{\mathrm{Re}\big(E(t)\exp[\mathrm{i}\{\omega_0 t + \phi(t)\}]\big) + U(t)\right\},$$

where ω_0 is the angular frequency corresponding to the 22-year period, $E(t)$ and $\phi(t)$ are the instantaneous envelope amplitude and phase, $U(t)$ is the Elatina undulation, and $\mathscr{H}$ is the nonlinear operator that puts in the third harmonic. To a good approximation, except in the neighbourhood of $R(t) = 0$, we have

$$R(t) = | R_{\pm}(t) | .$$

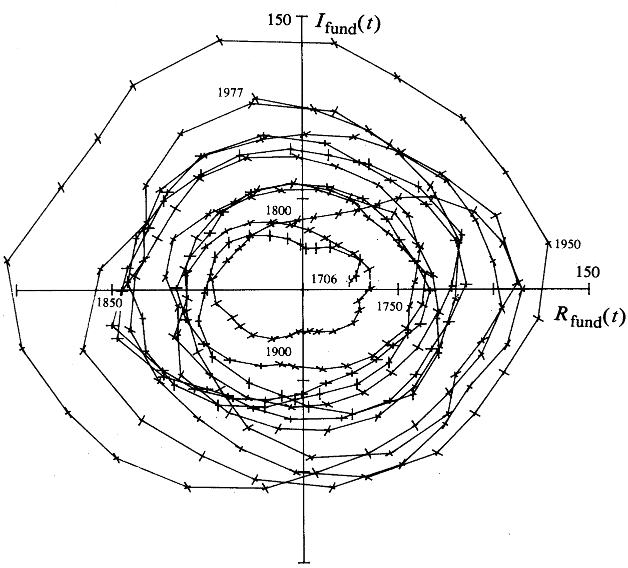

Fig. 7. Complex locus whose projection on the real axis is the series $R_{\mathrm{fund}}(t)$. The time marks are at yearly intervals.

9. Discussion

Envelope analysis. From the envelope analysis described above one can comment on physical questions such as the following: is the 22-year oscillation of the nature of quasi-monochromatic emission, as in an optical spectral line, or is it like a monochromatic oscillator whose signal arrives via a time-varying medium, as with ionospheric radio propagation links? The first case may be represented by a resonator tuned to a 22-year period and excited by random noise (cf. Yule 1926). The torsional oscillator invoked by Babcock (1961) [as distinct from the torsional motions observed by Howard and LaBonte (1980)] could be interpreted in these terms by supposing that the torsional vibration has a natural period of 22 years and characteristic damping and is excited by random mechanical motion or turbulence into a statistically steady state. An instance of the second case would be a submerged oscillator, or energy source, with much narrower line width than the natural bandwidth of the resonator, whose influence reaches the surface by multipath propagation through a turbulent medium. A third possibility would be a source that releases energy impulsively at intervals determined by a relaxation process, a mechanism that has been used in clocks, and that is currently entertained as Waldmeier's eruption hypothesis (Kiepenheuer 1953).

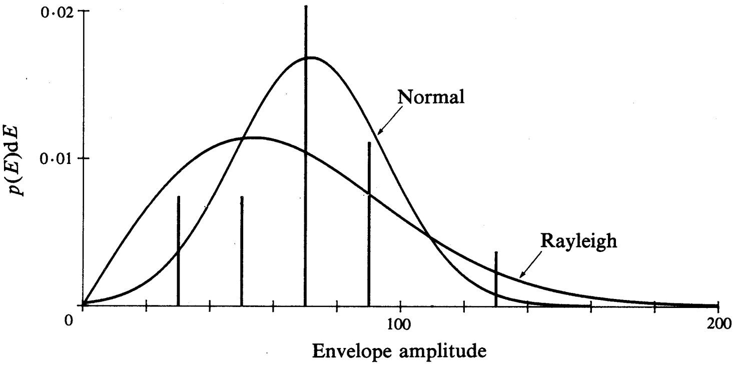

Fig. 8. Frequency distribution of envelope amplitude (histogram) which is reasonably matched by the normal distribution with mean 71·3 and standard deviation 23·7, but not by the Rayleigh distribution of best fit.

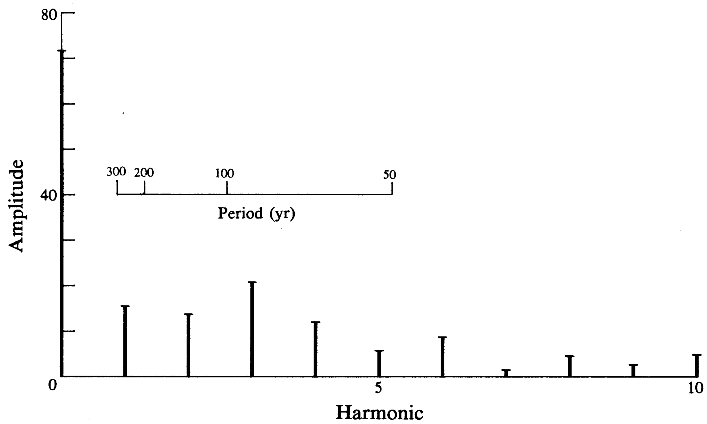

Fig. 9. Amplitudes of the coefficients of the Fourier series for the available envelope segment from 1706 to 1977.

Fig. 8 shows the frequency distribution $p(E)$ of envelope amplitude E defined so that $p(E)\,\mathrm{d}E$ is the frequency with which E is found between $E-\frac{1}{2}\mathrm{d}E$ and $E+\frac{1}{2}\mathrm{d}E$. Superimposed are respectively the best-fit normal distribution and Rayleigh distribution:

$$\left(\frac{1}{2\pi\langle E^2\rangle}\right)^{\frac{1}{2}} \exp\{-(E-\langle E\rangle)^2/2\langle E^2\rangle\} \quad \text{and} \quad \frac{2E}{\langle E^2\rangle}\exp\{-E^2/\langle E^2\rangle\}\,.$$

The comparison numerically favours the normal distribution over the Rayleigh distribution, as tested by the parameters of ratio of standard deviation to mean, skewness, and asymptotic fall-off, and thus weighs against the quasi-monochromatic behaviour which is known to exhibit a Rayleigh distribution of instantaneous amplitude. Confirmation of this result will be possible by appropriate analysis of the fossil record.

Fig. 9 shows the amplitudes of the Fourier series coefficients for the segment of envelope in Fig. 5. It has not been possible to perform this analysis before, in the absence of a method of extracting the envelope. Hartmann (1971) worked out upper and lower envelopes which were different from each other, but did not subject them to Fourier analysis, and concluded by averaging intervals between peaks, that 'one cycle lasts 73 years'. Now, one sees from the spectrum that the envelope variation is spread over several harmonics, with comparable but weak amplitudes. Only a trace of support for an 80-year cycle can be seen; it does not appear that the widely mentioned 80-year cycle is of any particular significance.

An interpretation of the amplitude modulation may be based on the model of a submerged monochromatic oscillator whose influence is transmitted to the observable solar surface through a time-varying medium. In this model, variations in both length and strength of the 11-year cycle would be attributable to a mechanism such that a r.m.s. envelope variation at the surface of about 0·4 of the mean is introduced by the inconstant medium. If the 'influence' is pictured in the concrete form of a rising wave of buoyant magnetic flux then phase advances and delays of the order of a year, and occasionally more [see the graph of $\Delta\phi(t)$ in Fig. 6], would have to be explained for the fluctuating component. Multipath propagation or indirect scatter are not indicated by the envelope statistics as a cause of amplitude variation; on the contrary, a mechanism for time-variable modulation of a single ray is suggested. Turbulence or other mechanical motion could vary the coupling between the oscillator and its propagation medium or attenuate or scatter the rising flux waves; a different specific mechanism is considered below. A time scale for change of the order of a century is indicated by the autocovariance function of the envelope. The location of the proposed monochromatic oscillator and the time-varying propagation medium may be rather deep because the length and strength of the northern and southern cycles do not disagree much by comparison with the variation from cycle to cycle. However, on the occasion of the large phase excursion of 1788, the two hemispheres may not have fully shared the perturbation. In that case the years of maximum sunspot number need not have been the same for both and the northern and southern minima may have been staggered, which would tend to raise the minimum count; there is an indication of this in the sunspot record, judging by the minima.

Phase analysis. We turn now to the analysis of phase. Dicke (1978) studied the solar cycle phase as specified by the occurrence of sunspot maxima and noticed that 'starting with the sunspot maxima of 1761·5 there occurred a remarkable series of three very short half-cycle periods, with an average length of 8·9 yr, after which sunspot maximum occurred $\sim$5·6 yr too early.' He then noted: 'But this was followed by a 17·1 yr half-cycle that completely corrected the phase error. It is as though the Sun 'remembered' the correct phase for 27 yr and then suddenly reset the sunspot cycle.' As one possible explanation, Dicke proposed that a strong cycle might transport its magnetic field to the surface sooner because of the greater buoyancy of the strong magnetic flux.

The phase anomaly referred to by Dicke is clearly visible on the graph of $\phi(t)$. It is the biggest variation of its kind in the record but other positive anomalies are noticeable (centred on 1846 and 1965), and negative anomalies (1827 and 1911). Also there are other strong cycles that can be tested for accompanying phase variations of the kind he proposed; to do this, refer to the phase departure $\Delta\phi$ with respect to the straight line in Fig. 6. In Fig. 10 $\Delta\phi(t)$ is plotted against the envelope amplitude

$E(t)$ for the years of maximum sunspot number. The correlation coefficient between $\Delta\phi - \langle\Delta\phi\rangle$ and $E - \langle E\rangle$ is $0{\cdot}59$ and thus tends to confirm the association of phase advance with strong cycles. The event culminating in 1787 stands out in the first quadrant. There were other large excursions of amplitude less obviously correlated with phase; but even if the 1787 event is deleted the correlation remains. The regression line shown has a slope $0{\cdot}0018$, corresponding to $\Delta t = 0{\cdot}02\{E(t) - 72\}$, where Δt is the relative time advance in years.

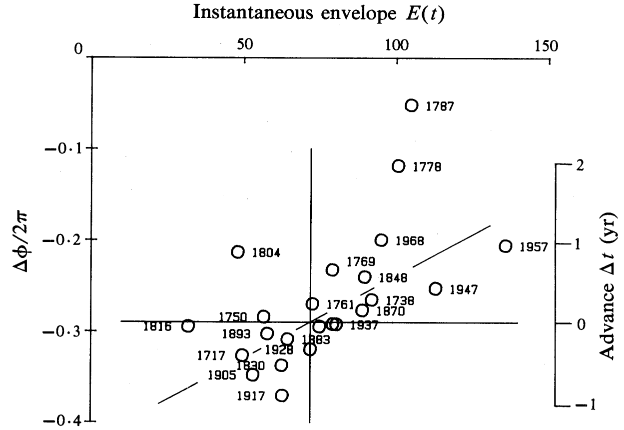

Fig. 10. Departure of instantaneous phase from the linear variation $26\pi(t-1700)/(1988-1700)$ against the envelope amplitude.

Effect of buoyancy. The three short cycles followed by a long one, or 'the remarkably large fluctuation in the intervals between sunspot maxima', cannot, however, be directly explained by the notion that transit time from interior to surface might be less in accordance as the amount of magnetic flux rising is greater. Such an acceleration would alter the waveform but would leave the cycle length unchanged. An analogous phenomenon is seen with water waves, where the velocity depends on depth and the higher wavecrests experience reduced transit times; the same is true of loud sound. In consequence the wavefront steepens. Rather than a shortened sunspot cycle length one would expect a faster rise from minimum to maximum to result from extra buoyancy in strong cycles. Indeed, in each of the three short cycles we find a rise time of only three years. It may be concluded that the transport interval as influenced by buoyancy could be the explanation of the solar cycle asymmetry. Then the correlation confirmed in Fig. 10 requires its own separate explanation.

Moving medium. To explain why the strong cycles are associated with a positive phase anomaly lasting for thirty years or so, we note that not only were there several short cycles but that the whole group of cycles arrived early. If the oscillator is monochromatic then upward convection of the propagation medium or some equivalent transit-time or path reduction mechanism would offer an explanation. Steady upward convection of the medium would not in itself shorten the cycle length seen at the surface; the magnetic field cycles would need to be compressed as they rose. It follows that the short cycles could have been brief because they were convected up in an era when they were being compressed in their direction of propagation.

Cycle asymmetry. The magnetic field being strengthened by compression, the extra buoyancy of the crests produced the steep fronts evidenced as short rise times, as noted above. The inverse consequence would be that in an era of path lengthening, the surface cycle length would be lengthened, cycles would be rarefied and weakened, and the waveform would be more symmetrical. These extremities of asymmetry and symmetry were in fact established by Waldmeier in connection with the eruption hypothesis; now they are explicable as consequences of propagation through a moving medium.

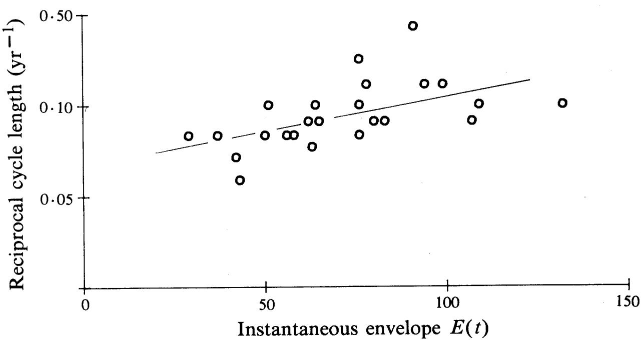

Fig. 11. Correlation between the reciprocal of cycle length and envelope amplitude showing that in times when the envelope $E(t)$ was large the cycle length was short, and vice versa, as predicted for propagation in a moving medium.

Correlation of cycle length with envelope. A further feature, not hitherto ascertainable because of the unavailability of a good envelope, but predictable on the basis of a moving medium, would be a correlation between the envelope amplitude and the reciprocal cycle length. Fig. 11, which shows the reciprocal cycle length between maxima against envelope amplitude, crisply confirms, with a correlation coefficient of 0·54, that this correlation exists. The change Δt in cycle length is given by $\Delta t = -0{\cdot}043\{E(t)-72\}$.

Doppler-shift interpretation. It is unnecessary to suppose that compression of successive half cycles of magnetic flux implies compression of the plasma. In order to account for the expansion and contraction of pathlength, which is all the present conception requires, the ray path need only be subject to time-varying refraction. If the path upward were not always vertical, but susceptible to sinuous deviations on a time scale of a century, then the variation of the pathlength would produce the effects seen. Under this interpretation the variability of the sunspot cycle length is describable as a Doppler shift associated with variation of the phase pathlength between oscillator and surface. The Wolf modulation of envelope amplitude is an associated consequence; an analogous phenomenon exists in fixed point to fixed point transmission of sound waves through the atmosphere in the presence of moving temperature gradients.

The correlation of cycle length with envelope presented in Fig. 11 is predicted by the Doppler-shift interpretation since, in the presence of flux conservation, compression of the cycle length will strengthen the magnetic field.

Interconnection of phase derivative with amplitude. If a ray path from source to observable surface were in slow motion, the phase advance $\Delta\phi(t)$ would normally be negative because the average ray would be longer than the shortest possible ray. On an occasion when the ray length passed through a minimum not much different from the shortest possible, there would be a rising and falling phase anomaly, played out on a short time scale. On the rising slope of the anomaly (diminishing ray pathlength) there would be positive Doppler shift (shortening of the cycle length) and the opposite on the falling slope. An amplitude increase associated with compression of the cycles would tend to peak at times of maximum Doppler shift which are the same as times of maximum rate of change $\Delta\phi'(t)$. Therefore, it is of interest to look into the possibility that the derivative of the instantaneous phase is connected with the instantaneous amplitude of the envelope. This is done in Fig. 12, where the envelope $E(t)$ is the solid curve, and the dashed curve is the derivative, represented by the first difference of $\Delta\phi(t)$ over an interval of 10 yr plotted at 5-year intervals. There is indeed an interdependence between the two curves but the agreement would be much better if the derivative were delayed 10 yr. This discrepancy might be connected with ray convergence as a factor contributing to amplitude increase, with properties of the unknown dispersion relation for internally propagated magnetic flux waves, or with boundary conditions at the surface, none of which was invoked in the reasoning that led to the carrying out of the comparison. Meanwhile the present work establishes that the Wolf modulation is not independent of the variation in cycle length; but let us suppose that dispersion is responsible for the desynchronization.

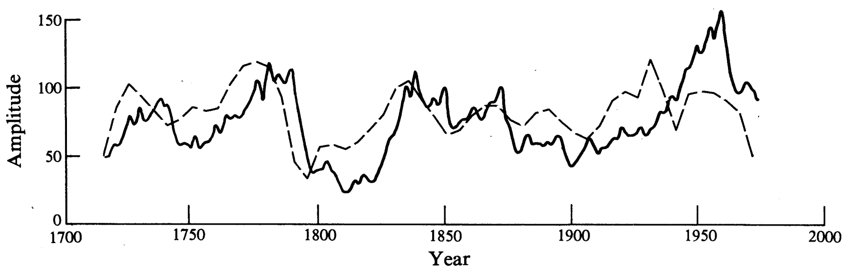

Fig. 12. Instantaneous envelope $E(t)$ (solid curve) compared with the derivative of the instantaneous phase (dashed curve).

Dispersive magnetic flux wave. Let $R_{\text{fund}}(s, t)$ be a field variable, at ray coordinate s, which at the surface is equal to the variable $R_{\text{fund}}(t)$ introduced in Section 6. A simple nonlinear, dispersive differential equation that produces the wavefront steepening and dispersion is

$$\frac{\partial^2 R_{\text{fund}}}{\partial s^2} = \frac{1}{c^2(1+kR_{\text{fund}}^2)}\frac{\partial^2 R_{\text{fund}}}{\partial t^2} + \epsilon R_{\text{fund}},$$

where c is the velocity of propagation of magnetic flux waves of infinitesimal amplitude, ϵ is a measure of dispersion, and k is the compliance per unit length of the propagation medium in units appropriate to the conversion ratio between wave disturbance and R_{fund}. From the properties of such an equation it may be deduced that $(v_{\text{ph}} - v_{\text{gr}})/c$, the fractional excess of phase over group velocity, is $\epsilon c^2/\omega^2$. If this excess velocity is to produce a 10-year lead then some connections follow between transit time, ray pathlength and dispersion. This type of differential equation implies a cutoff frequency ω_c such that $\omega_c^2 = \omega^2(v_{ph} - v_{gr})/c$.

Explanation of Elatina undulation. The form of an oscillation at the frequency ω_c has the character of a magnetic flux wave of infinite wavelength; the magnetic field would be parallel to the field of the 22-year flux wave, enhancing and counteracting it in alternate 11-year cycles, in exactly the way needed to produce the zig–zag effect. A requisite topological feature of whatever causes the Elatina undulation is that at a time when the cycle is enhanced in one hemisphere it must also be enhanced in the other hemisphere which, because of the opposition of sunspot group polarities, means that the cause, if additive, must reverse in sign between hemispheres. This is not an easy topology to account for but would be an attribute of the infinite-wavelength mode which, because of the boundary condition in the equatorial plane, would need to have an antinode there. Assuming that the Elatina undulation is indeed due to oscillation in the infinite-wavelength mode, one finds that

$$(v_{ph} - v_{gr})/c = \omega_c^2/\omega^2 = 1/30^2,$$

where the factor 30 is $T_U/11$. It follows that there are 450 waves in transit in the dispersive pipeline to account for a 10-year separation. Also the dispersion constant ϵ is constrained by $\epsilon c^2/\omega^2 = 1/30^2$.

10. Conclusions

The concept of a deep monochromatic oscillator launching 22-year magnetic flux waves through a time-varying propagation medium to influence the observable surface has been generous in supplying ideas for explaining several features of the sunspot number series. To go further will mean identifying the wave mode that is compatible with the parameters T_U, ϵ, c and the fluctuation level of one-seventh. Support for this investigation can be expected from statistical observations of surface magnetic fields and motions and from a knowledge of interior conditions revealed by helioseismology.

There is an extensive literature on the analysis of sunspot number series but the investigations have not hitherto led to results bearing on the physics of the solar cycle. Likewise, the part of solar physics concerned with the solar cycle has been largely observational and descriptive and, as regards theory, conjectural. Conclusions relating to the physical nature of the source of the oscillations have been practically nonexistent. Consequently, it is encouraging to see that the sunspot number series can contribute to the fundamental physical question. The rich fossil record unearthed by Williams (1981, 1983) will provide indispensable data for future progress on the physics of the solar cycle.

Acknowledgments

The bulk of the work reported here was performed at the CSIRO Division of Radiophysics while the author was on sabbatical leave from Stanford University. I thank the Chief of the Division, Dr R. H. Frater, for photographic and other facilities and for lending me his HP85 computer with which I did the data analysis and drew the figures. I also thank my colleagues A. C. Fraser-Smith, R. A. Helliwell and L. R. O. Storey for their comments.

References

Babcock, H. W. (1961). *Astrophys. J.* **133**, 572–87.
Born, M., and Wolf, E. (1959). 'Principles of Optics' (Pergamon: New York).
Bracewell, R. N. (1953). *Nature* **171**, 649–50.
Bracewell, R. N. (1978). 'The Fourier Transform and Its Applications' (McGraw-Hill: New York).
Dicke, R. H. (1978). *Nature* **276**, 676–80.
Gabor, D. (1946). *J. Inst. Electr. Eng.* **93**, 429.
Giovanelli, R. G. (1984). 'Secrets of the Sun' (Cambridge Univ. Press).
Hartmann, R. (1971). *Sol. Phys.* **21**, 246–8.
Howard, R., and LaBonte, B. J. (1980). *Astrophys. J.* **239**, 133–6.
Kiepenheuer, K. O. (1953). *In* 'The Sun' (Ed. G. P. Kuiper), Ch. 6 (Chicago Univ. Press).
Sonett, C. P. (1984). *Nature* **306**, 670–3.
Williams, G. E. (1981). *Nature* **291**, 624–8.
Williams, G. E. (1983). *In* 'Weather and Climate Responses to Solar Variations' (Ed. B. M. Cormac), pp. 517–33 (Colorado Assoc. Univ. Press).
Wolf, R. (1862). *Astronomische Mitteilungen Zürich*, No. 14.
Yule, G. U. (1926). *Phil. Trans. R. Soc. London* A **226**, 267.

Manuscript received 11 March, accepted 12 August 1985