Our changing Sun

Article

Published Version

Our changing Sun

M Lockwood, R Stamper, M N Wild, A Balogh and **G Jones** discuss how new research has been applied to historical data to improve our understanding of the Sun and its effect on Earth.

Stellar astronomy tells us much about the long-term evolution of our Sun while forensic evidence (for example, cosmic-ray products in ice cores) gives us indications of its fluctuations over the last millennium. However, such studies do not give us a sufficiently detailed understanding of solar change over the last century to allow us to detect and quantify any role that the Sun might have played in the observed rise in average surface temperatures on Earth. This paper describes recent research that has filled this gap by applying advances in our understanding of the effects and structure of the solar wind to historical data on the Earth's magnetic field.

A continuous sequence of data on geomagnetic activity extends back to 1868 (covering sunspot cycles 11–22) and shows a consistent rise since the turn of the century. Near-Earth interplanetary space has been routinely monitored since 1964, giving data throughout the last three of these sunspot cycles (20–22). Recently, it has been shown that the largest of three major contributions to the rise over these last three cycles is an upward drift in the magnitude of the interplanetary magnetic field (IMF). In addition, the Ulysses spacecraft has made the first out-of-ecliptic observations and shown that the latitudinal gradients in the radial component of the IMF are small. By combining these results, we can deduce that the total "source" magnetic flux in the Sun's atmosphere has risen by 41% since 1964. Using three extremely significant and theory-based correlations derived from data for solar cycles 21 and 22, we can estimate the coronal source flux using only the geomagnetic data. The method has been subjected to an independent test using IMF data for cycle 20 and the estimated mean value for the cycle shown to be accurate to within a few percent. This allows us to extrapolate the coronal field estimates back to 1868 with confidence: we find an increase by 130% since 1901. Furthermore, this coronal source flux has been shown to be very highly correlated with the Sun's brightness, giving a rise in the average power output of the Sun of $\Delta I = 1.65 \pm 0.23\ \mathrm{W\,m^{-2}}$ during this century. Such a large long-term change has important consequences for global warming. In this paper we first review the evidence for these changes and then discuss their implications.

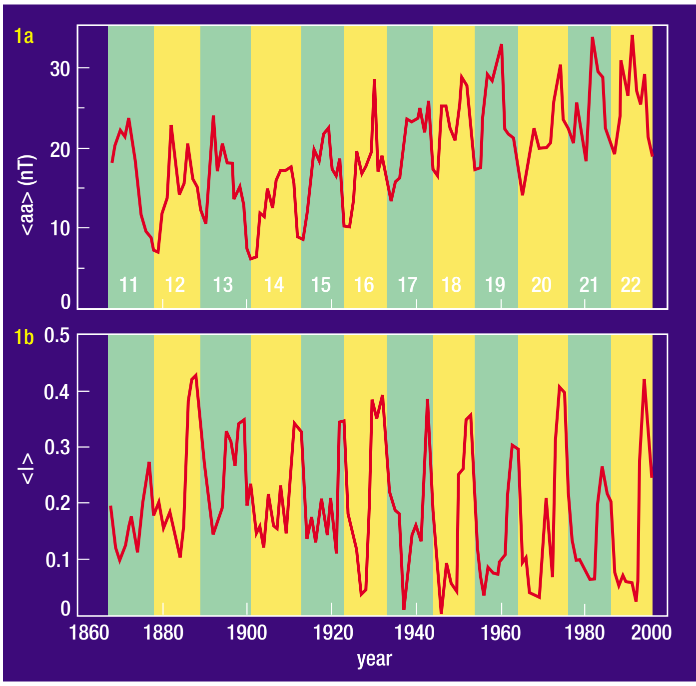
Geomagnetic activity

The *aa* index was devised by Mayaud (1972) to quantify the fluctuation level of the geomagnetic field. It employs the maximum range of variation in the field at the Earth's surface during three-hourly intervals, recorded since 1868 by pairs of near-antipodal magnetometers in England and Australia. Because the instruments used have been carefully cross-calibrated and because the data have been processed in a uniform way, this is a highly valuable and homogeneous data series. Figure 1a reveals that the annual means <*aa*> show a marked solar cycle variation. The other major feature of this long data series is the gradual drift to larger values. Recently, Clilverd *et al.* (1998) have analysed the potential causes of this increase in geomagnetic activity and Stamper *et al.* (1999) have shown that it is almost entirely due to changes in the interplanetary medium surrounding the Earth.

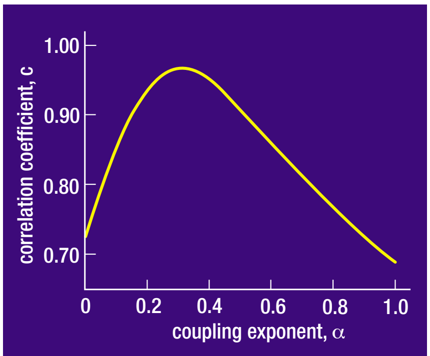
Geomagnetic activity, as quantified by indices such as *aa*, is caused by the energy of the solar wind. Because it is a large-scale plasma of high electrical conductivity, the "frozen-in flux theorem" applies to the solar wind which means that it draws magnetic flux out of the Sun and into the heliosphere. This field is the IMF which, through the process of magnetic reconnection, controls how much of the solar-wind energy is extracted by Earth's magnetosphere (see Lockwood 1997). The solar wind and IMF have been routinely monitored since the start of the space age and we now have three full solar cycles of interplanetary data to compare with the *aa* data. Short-term increases in *aa* are caused by transient solar disturbances, such as coronal mass ejections, hitting the Earth's magnetosphere. These events are more frequent at sunspot maximum (Webb and Howard 1994). In addition, and particularly during the declining phase of each sunspot cycle, *aa* can rise because Earth intersects long-lived, fast solar-wind streams (Cliver *et al.* 1996). These emanate from coronal holes that have expanded towards low heliographic latitudes and rotate with the equatorial photosphere every 25 days (Wang *et al.* 1996). During this time, the Earth moves along its orbit, such that it intersects the same stream every 27 days, giving recurrent geomagnetic activity that is quantified by an index *I*, derived from the autocorrelation function of *aa* at a lag of 27 days (Sargent 1986). Figure 1b shows that *I* tends to show strongest peaks in the declining phase of even-numbered sunspot cycles, as does the mean solar wind velocity, v_{sw} (Hapgood 1993; Cliver *et al.* 1996).

Several attempts have been made to use geomagnetic data to deduce the interplanetary and solar conditions before space measurements began (Russell 1975). The success of such an extrapolation depends on the quality of the correlation found between the geomagnetic index and the combination of the interplanetary parameters (the empirical "coupling function") used to quantify the controlling influence of the solar wind and IMF (Baker 1986). An early attempt at extrapolation used data from solar cycle 20 only (Gringauz 1981) and was based on a correlation between *aa* and v_{sw}. As data for solar cycle 21 were received, it became clear that a much better correlation was obtained if a dependence on the southward component of the IMF was also included (Crooker and Gringauz 1993). This was used to look at the possible combinations of v_{sw} and the IMF that existed at the turn of the century (Feynman and Crooker 1978). Recently, Stamper *et al.* (1999) obtained an unprecedentedly high correlation coefficient of 0.97 (using a coupling function that is a theory-based combination of v_{sw}, the IMF magnitude B_{sw}, the IMF orientation, and the solar wind concentration N_{sw}), whereas the correlations for all previously proposed coupling functions were degraded by the addition of data for solar cycle 22. The coupling function derived by Stamper *et al.* formed the basis of an extrapolation back to 1868 by Lockwood *et al.* (1999a).

Lockwood *et al.* (1999a) developed a novel method for estimating the IMF magnitude B_{sw}



1: Annual means of (a) the *aa* geomagnetic activity index, <*aa*> and (b) Sargent's 27-day recurrence index <*I*>. The data are for 1868–1996, covering sunspot cycles 11–22: the solar cycles, commencing at sunspot minimum, are shaded alternately yellow and green. For all phases of the solar cycle, <*aa*> increased gradually between 1900 and about 1955, before decreasing and then rising again during 1964–1996.



2: The correlation coefficient, *c*, between annual means of the power input to the magnetosphere P_α and *aa*, as a function of the coupling exponent α. P_α is derived from interplanetary measurements using equation (1), with data from solar cycles 21 and 22. An unprecedentedly high correlation of 0.966 is obtained at $\alpha = 0.3085$.

from the *aa* data. This exploits two strong and extremely significant correlations between the IMF, the solar wind and the *aa* index, which Lockwood *et al.* derived using the data from the last three solar cycles (20–22). However, there are uncertainties concerning the calibration of the early interplanetary measurements (Gazis 1996), particularly for N_{sw} in solar cycle 20. Consequently, Lockwood and Stamper (1999) employed a different approach. They derived all correlations using data from cycles 21 and 22 only and then predictions for cycle 20 were compared with the IMF observations. Thus the cycle-20 IMF data provided an independent test of the method.

The method makes use of the theory of energy transfer into the Earth's magnetosphere by Vasyliunas *et al.* (1982). The power delivered to the magnetosphere, P_α, is the multiplied product of three terms: (1) the energy flux density of the interplanetary medium surrounding the Earth (dominated by the kinetic energy of bulk solar-wind flow); (2) the area of the target presented by the geomagnetic field (roughly circular with radius l_o); (3) the fraction t_r of the incident energy that is extracted. The dayside magnetosphere is approximately hemispherical in shape, in which case l_o equals the stand-off distance of the nose of the magnetosphere which can, to first order, be computed from pressure balance between the Earth's dipole field and the solar-wind dynamic pressure. We adopt the form of the dimensionless "transfer function" t_r, as suggested by Vasyliunas *et al.*, which includes an empirical $\sin^4(\theta/2)$ dependence on the IMF clock angle θ (the angle that the IMF makes with northward in the GSM frame of reference) (Scurry and Russell 1991) and thereby allows for the role of magnetic reconnection between the IMF and the geomagnetic field. The transfer function adopted also depends on the solar wind Alfvén Mach number, to the power 2α, where α is called the "coupling exponent" and must be determined empirically. This yields (Stamper *et al.* 1999):

$$P_\alpha = k m_{sw}^{(2/3-\alpha)} M_E^{2/3} B_{sw}^{2\alpha}\left[N_{sw}^{(2/3-\alpha)} v_{sw}^{(7/3-2\alpha)} \sin^4(\theta/2)\right] = k m_{sw}^{(2/3-\alpha)} M_E^{2/3} B_{sw}^{2\alpha} f \quad (1)$$

where m_{sw} is the mean ion mass of the solar wind, M_E is the magnetic moment of the Earth and k is a constant. To compute α, the *aa* index is assumed to be proportional to the extracted power $P_\alpha = (aa/s_a)$, an assumption that is verified by the scatter plot shown in figure 3a. The optimum α gives the peak correlation coefficient c between P_α and *aa* (see figure 2). The constant $s_a' = ks_a$ is then found from a linear regression fit (lower panel, figure 3a).

Stamper *et al.* (1999) have analysed each of the terms in the best-fit coupling function given by equation (1). They showed that more than half of the change in *aa* over the last three solar cycles was caused by an upward drift in B_{sw}. There were also contributions from increases in N_{sw} and v_{sw} but the average IMF clock angle θ had grown slightly less favourable for causing geomagnetic activity (because there was a slight tendency for the IMF vector to stay closer to the ecliptic plane).

In order to use equation (1) to evaluate B_{sw}, we group the terms in the square brackets together into a parameter f, the variation of which (on annual timescales) is dominated by that in v_{sw}. The annual mean of v_{sw} rises in the declining phase of solar cycles (particularly even-numbered ones) (Cliver *et al.* 1996; Hapgood 1993) because the Earth then repeatedly intersects the fast, low-density solar-wind streams that emanate from the low-latitude extension of coronal holes. These multiple intersections occur every 27 days and so also raise the recurrence index I shown in figure 1b. Hence both f and I increase together in the declining phase of sunspot cycles. However, I tends to remain high towards sunspot minimum, whereas v_{sw} and f are lower, because *aa* values are low and relatively constant. Consequently, Lockwood *et al.* (1999a) adopted a relationship for a predicted f of the form:

$$f_p = s_f I^\beta aa^\lambda + c_f \quad (2)$$

where the exponents β and λ give the optimum correlation coefficient and the constants s_f and c_f are then found from a linear regression fit of f against f_p (figure 3b). Substituting for f in (1), using f_p given by (2), allowed Lockwood *et al.* to compute B_{sw} from the *aa* data series. They employed estimates of M_E from the IGRF model fit to geomagnetic data and assumed the composition of the solar wind is constant with a mean ion mass of 1.15 a.m.u.

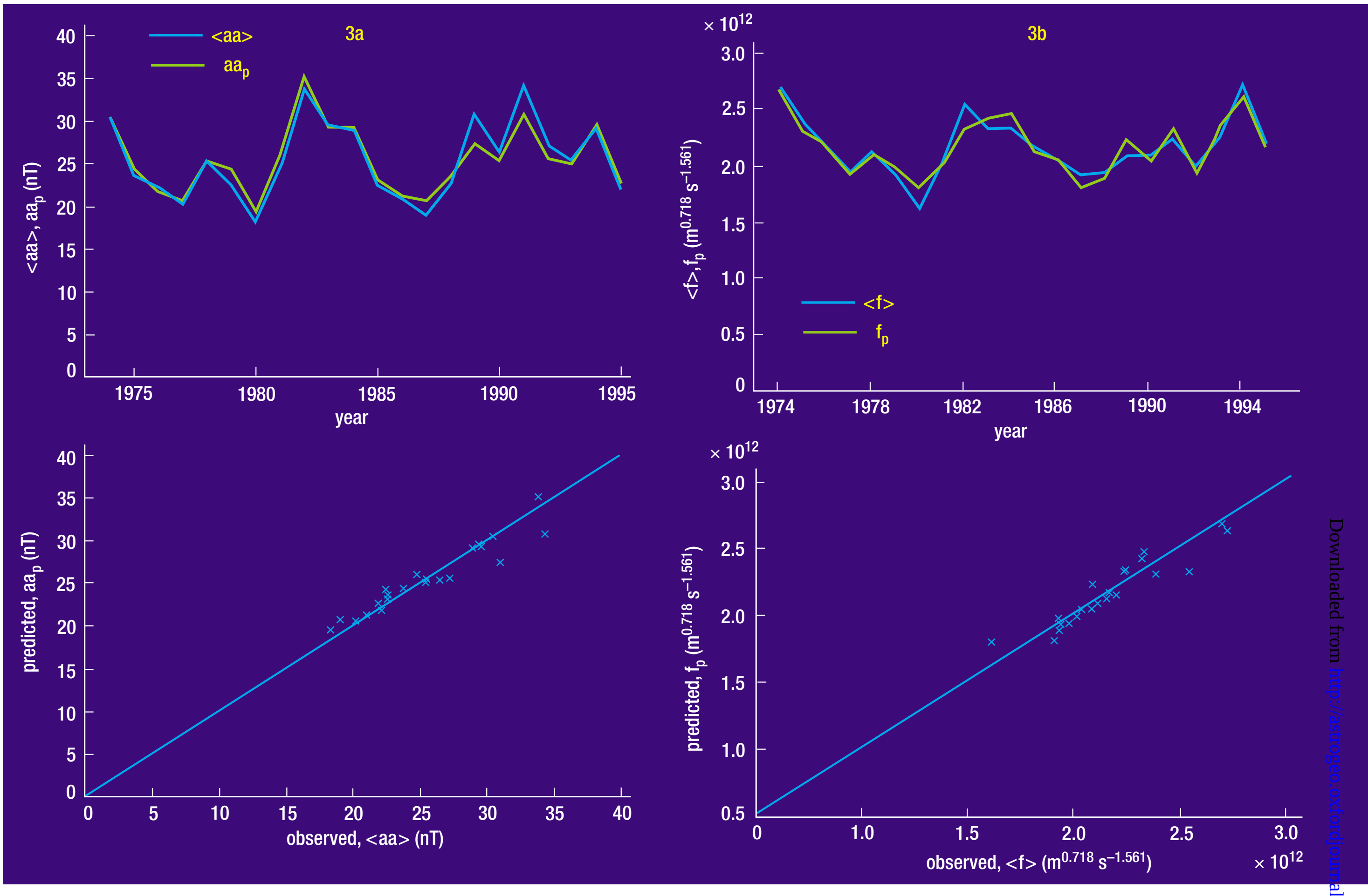
The heliospheric field

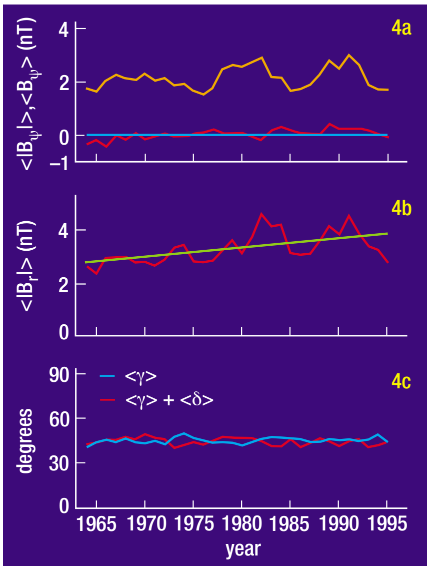
Parker spiral theory (e.g. Gazis 1996) predicts the heliospheric field in heliocentric polar coordinates (r, ϕ, ψ) will be:

$$B_{sw} = \{B_r^2 + B_\phi^2 + B_\psi^2\}^{1/2} = B_r\{1+\tan^2\gamma\}^{1/2} = B_o(R_o/r)^2\{1+(\omega r\cos\psi/v_{sw})^2\}^{1/2} \quad (3)$$

where B_o is the coronal source field at the solar source sphere, $r = R_o \approx 2.5R_s$ from the centre of the Sun, where the solar field becomes approximately radial ($B_o \approx B_r$) (Wang and Sheeley 1995); ω is the equatorial angular solar rotation velocity, ψ is the heliographic latitude and R_s is the solar radius.

Figure 4 demonstrates that Parker spiral theory is very successful in predicting annual means of the heliospheric field orientation around Earth (Stamper *et al.* 1999) because phenomena causing short-term perturbations, such as corotating interaction regions (CIRs) and transient coronal mass ejections (CMEs),



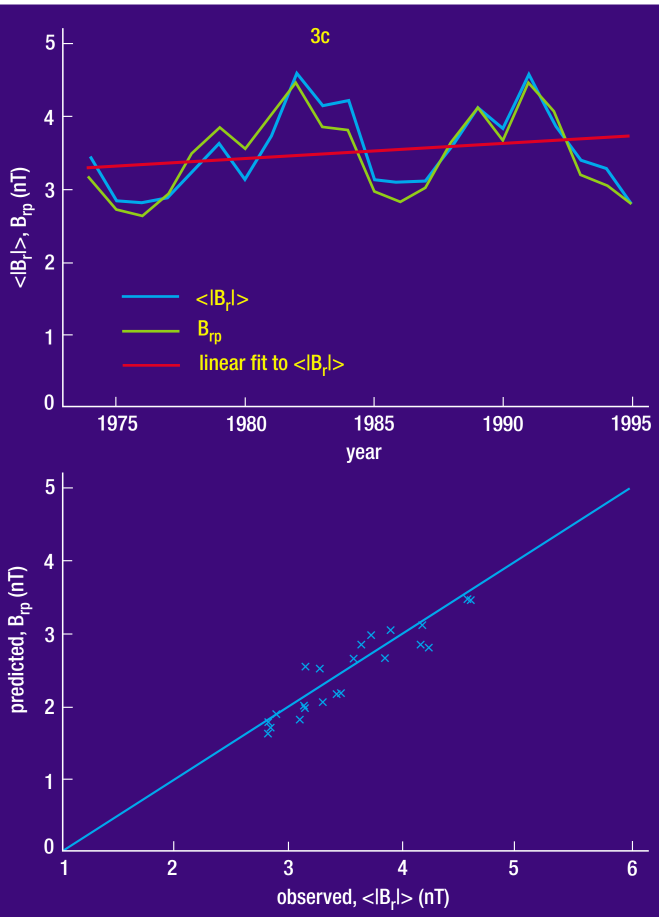


4: Annual means of interplanetary magnetic field components for cycles 20–22 (in red) compared to the predictions of Parker spiral theory (in blue). (a) The mean of the modulus of the out-of-ecliptic IMF component $<|B_\psi|>$ (orange line) and the mean $<B_\psi>$ (red line) which remains close to zero (blue line). (b) The modulus of the radial component of the observed IMF $|B_r|$, the green line is a linear fit to the three full cycles. (c) shows that the "garden hose angle" γ of the IMF in the ecliptic plane (equal to $\tan^{-1}B_\phi/B_r$): both the observed angle (red line) and that predicted by equation (3) (blue line) remain close to 45° in these annual means (after Stamper *et al.* 1999).

are averaged out. Figure 4a shows the annual means of the modulus of the out-of-ecliptic IMF component $<|B_\psi|>$ (in orange), which is well-correlated with B_{sw}, and the mean $<B_\psi>$ (in red), which remains close to zero as is predicted by the theory. Figure 4c shows the "garden hose angle" γ of the IMF in the ecliptic plane (equal to $\tan^{-1}B_\phi/B_r$): both the observed angle (red line) and that predicted by equation (3) (blue line) remain close to 45° and, as a result, the radial heliospheric field component $|B_r|$ is roughly proportional to B_{sw}, i.e. $|B_r| \approx |B_{rp}| = s_B B_{sw}$ (see figure 3c).

Figure 4b shows the annual means of the radial component of the observed IMF. It reveals an upward drift superposed on the solar cycle variation. The validity of equation (3) on these annual time scales, revealed by figure 4, tells us that these variations in $|B_r|$ reflect variations in the coronal source field, $|B_o|$. The green line is a linear regression fit and reveals that the increase is by a factor of 1.3 over these three full solar cycles. With better allowance for magnetogram saturation effects, Wang and Sheeley (1995) modelled the coronal field from measurements in the underlying photosphere, using the discovery by the Ulysses spacecraft that there are sheet currents, but no significant volume currents, in the heliosphere (Balogh *et al.* 1995). Although they did not comment on it, their results do also show this upward drift.

The knowledge gained by the Ulysses spacecraft is vital. It has given us our first view of the heliosphere from out of the ecliptic plane. From its initial journey from the ecliptic to above the Sun's southern pole, Balogh *et al.* (1995) found that the latitudinal gradients in $|B_r|$ are small. This result was confirmed by the fast pole-to-pole solar pass by the spacecraft between 13 September 1994 and 31 July 1995, as demonstrated here in figure 5. During this pass, the $(r/R_1, \psi)$ co-ordinates of Ulysses vary from (2.29, –80.2°) to (2.02, +80°), crossing the heliographic equator ($\psi = 0$) on 4 March and reaching perihelion at (1.34, +6°) on 12 March (where $R_1 = 1$ AU). The data have been averaged into 26 day periods (roughly mean solar rotation periods) to remove longitudinal solar structure, CIRs and CMEs. The orange line shows the radial field, normalized to the Earth's orbit using the r^2 variation predicted by equation (3), $B_r(r/R_1)^2$. The data show the inward field ($B_r < 0$) in the southern heliographic hemisphere and the outward in the northern ($B_r > 0$). The average B_r close to the heliographic equator is near zero because the current sheet was crossed several times. The blue line shows the modulus of the normalized radial field, $|B_r|(r/R_1)^2$, which shows very little latitudinal structure. The green line shows the average value for the pass, $<|B_r|(r/R_1)^2>$ which agrees with the value seen when the spacecraft was closest to the ecliptic plane to within an uncertainty of 2.7%. This close agreement means that the



3: Three correlations used to compute the solar source flux from the *aa* index, derived from annual means of data for solar cycles 21 and 22. In each part, the top panel shows the time series of observed and best-fit values, and the lower panel shows the same data as a scatter plot. Parameters, the best-fit exponents and the linear regression coefficients are all given in table 1. (a) the observed *aa* index *aa* and the value predicted from interplanetary measurements aa_p; (b) the observed interplanetary parameter f and the value predicted from aa and l, f_p; and (c) the observed modulus of the radial component of the IMF, $|B_r|$ and its value predicted from the IMF magnitude, $B_{rp} = s_B B_{sw}$. The red line in (c) is a linear regression fit to $|B_r|$ for the solar cycles 21 and 22 and shows that the rise over cycles 20–22 (see figure 4b) is not just due to the low IMF values in cycle 20.

Table 1: Regression fits used to extrapolate F_s

fitted parameters	correlation coefficient c	significance level (%)	coefficients	slope	intercept
aa and aa_p	0.966	$100-3.442\times10^{-11}$	$\alpha=0.3085$	$s_a'=4.7022\times10^{-18}$	–
f and f_p	0.934	$100-3.925\times10^{-8}$	$\beta=0.2271$ $\lambda=1.2114$	$s_f=5.71\times10^{5}$	$c_f=2.61\times10^{7}$
$\langle\lvert B_r\rvert\rangle$ and B_{rp}	0.928	$100-4.382\times10^{-8}$	–	$s_B=0.5606$	–

units:

aa_p (in nT) $= s_a' \langle M_E \text{ in Tm}^3\rangle^{2/3}\, m_{sw}^{(2/3-\alpha)} \langle N_{sw} \text{ in m}^{-3}\rangle^{(2/3-\alpha)} \langle v_{sw} \text{ in km s}^{-1}\rangle^{(7/3-2\alpha)} \langle B_{sw} \text{ in nT}\rangle^{2\alpha} \langle \sin^4(\theta/2)\rangle$

$f = \langle N_{sw} \text{ in m}^{-3}\rangle^{(2/3-\alpha)} \langle v_{sw} \text{ in km s}^{-1}\rangle^{(7/3-2\alpha)} \langle \sin^4(\theta/2)\rangle$ and $f_p = s_f \langle l\rangle^{\beta} \langle aa \text{ in nT}\rangle^{\lambda} + c_f$

B_{rp} (in nT) $= s_B \langle B_{sw} \text{ in nT}\rangle$

total flux threading the sphere of radius R_1 is, to a good approximation, given by:

$$F_s = (1/2)\, 4\pi R_1^2 |B_r| \approx 2\pi R_O^2 |B_O| \approx 2\pi R_1^2 s_B B_{sw} \quad (4)$$

This result is vital as it allows us to use the mean radial field in one place (i.e. near Earth) to calculate the coronal source flux, F_s – the total magnetic flux leaving the Sun and entering the heliosphere. (The factor of a half arises because half the flux is outward, half inward.)

The long-term variation of the coronal source flux

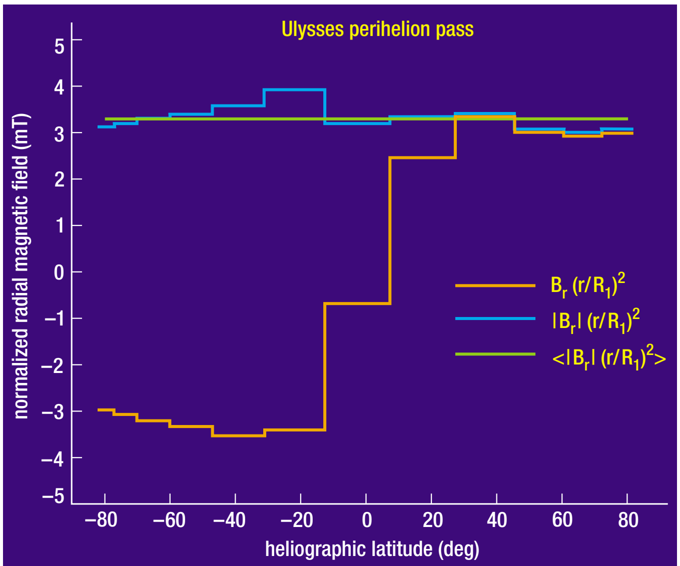
Equations (1), (2) and (4) give a method for computing F_s from the *aa* index, once the various exponents and coefficients have been derived from the data for cycles 21 and 22. Table 1 shows that the three correlations shown in figure 3 are all strong (correlation coefficients approaching unity) and extremely significant (almost zero probability of obtaining the result by chance). The least strong of the three correlations is that between the IMF magnitude and its radial component and this is responsible for most of the small error in the method for computing the coronal source flux F_s. The table also lists all the exponents and coefficients required.

Figure 6 shows the values of the coronal source flux. Those derived from *aa*, F_{sp}, are shown by the area shaded green, whereas those from near-Earth measurements of the IMF, F_s, are shown by the red line. The area shaded orange shows the variation of the smoothed sunspot number R for comparison. Because the data for cycle 20 are not included in the derivation, they provide a fully independent test. Table 2 gives the results of this test, as carried out by Lockwood and Stamper (1999). It shows the RMS deviation of F_{sp} from F_s is similar for all three cycles, so the method has reproduced the variation in annual means well – despite the fact that cycle 20 is very unusual and different from cycles 21 and 22 in many ways. Table 2 also gives the minimum-to-minimum averages, $\langle F_{sp}\rangle$ and $\langle F_s\rangle$. It can be seen that the error ε in $\langle F_{sp}\rangle$ is only 1.5% for the test data, actually rather better than the 4.5% for one of the fitted cycles. Thus the method has successfully extrapolated from cycles 21 and 22 to cycle 20. This means we can apply the method to all the *aa* data, back to 1868, with considerable confidence. Note that figure 6 is not significantly different from the results of Lockwood *et al.* (1999a) who used correlations based on data from all three solar cycles.

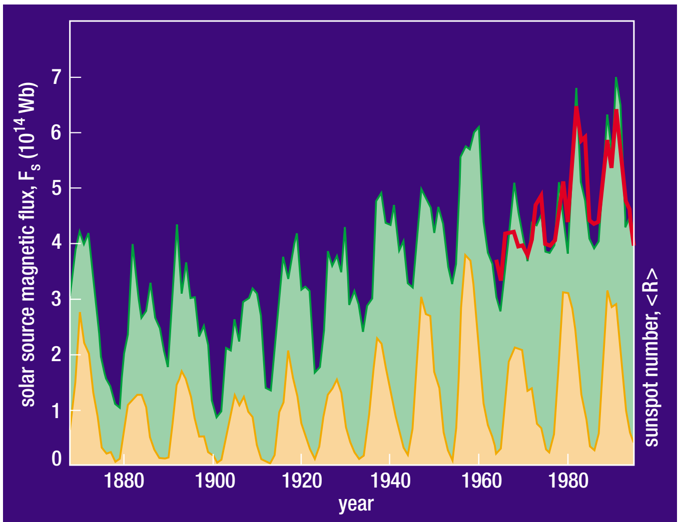
Figure 6 shows that the solar coronal source flux has a marked solar cycle variation, peaking shortly after sunspot maximum at about the time that the Sun's field changes polarity. Also, it has risen steadily since the turn of the century, with the exception of a fall in the 1960s. Figure 7 plots the minimum-to-minimum cycle means, $\langle F_{sp}\rangle$ (in orange) and $\langle F_s\rangle$ (in blue): the variation of $\langle F_{sp}\rangle$ is similar to that in the mean sunspot numbers $\langle R\rangle$ (in green). The change in F_{sp} also correlates well with changes in the mean length of sunspot cycles (Friis-Christensen and Lassen 1991) and in the mean latitude of sunspots (Pulkkinen *et al.* 1999). It is also consistent with changes in cosmogenic isotopes found, for example, in polar ice cores (Lean *et al.* 1995) (see the following section). The increase in the coronal source flux since cycle 14 is by a remarkably large factor of 2.31 (a rise of 131%).

Implications of the change

The change revealed by figures 6 and 7 is significant. A wide variety of operational systems are influenced by F_s through the power extracted by the magnetosphere from the solar wind. They include: communications, broadcast and radar systems; power distribution networks; spacecraft; oil pipelines; oil exploration rigs and radio and magnetic navigation and guidance systems. For example, the rise in F_s may mean that near-Earth space would have been a less hostile environment for satellites back in 1900 than it is now (Lockwood *et al.* 1999b) and future changes in F_s will be relevant to the design and safe operation of satellites. The upper atmosphere is known to be heated by the deposition of extracted solar wind energy and this heating will have risen with F_s and *aa*. However, the most intriguing, and controversial, possibilities are that the changes in F_s will be associated with changes in the Earth's lower atmosphere, oceans and climate.



5: Observations of the radial component of the IMF by the Ulysses spacecraft during its fast solar latitude scan about perihelion. The data have been averaged into 26-day periods. The orange line shows the radial field, normalized to the Earth's orbit, $B_r\,(r/R_1)^2$, where R_1 is 1 AU. The blue line shows the modulus of the normalized radial field, $|B_r|(r/R_1)^2$, which shows very little latitudinal gradient, and the green line shows the average value for the pass $<|B_r|(r/R_1)^2>$.



6: Annual means of the coronal source magnetic flux. Those derived from *aa*, F_{sp}, are shown by the area shaded green, whereas those from near-Earth measurements of the IMF, F_s, are shown by the red line. The area shaded orange shows the variation of the smoothed sunspot number. This plot is based on the correlations shown in figure 3, obtained using data from solar cycles 21 and 22 only (after Lockwood and Stamper, 1999).

The concept that significant climate change is caused by changes in the solar output is certainly not new (e.g. Blanford 1891); indeed, as early as 1801 Herschel argued that this effect was the cause of an apparent anticorrelation between sunspot numbers and wheat prices (Herschel 1801). Monitoring over the last 15 years has shown that the total solar irradiance I does indeed vary over the sunspot cycle (Willson 1997). There are two main contributions: firstly, sunspots being cooler, darker regions of the photosphere give a slight decrease in solar luminosity; however, this is outweighed by the positive contribution of brightenings like faculae which are associated with sunspots (Lean *et al.* 1995). These phenomena are both magnetic in origin, although they are near $r=R_s$ and in the region of closed solar flux, whereas F_s is mainly open flux near $r=2.5R_s$. However, if the change in F_s reflects a fundamental variation in the solar dynamo, we would expect a relationship between I and F_s. The observed solar cycle variation is small, I varying between 1367.0 W m^{-2} and 1368.3 W m^{-2}, a variation of just 0.1%. The various total solar output monitors are not well calibrated, but Lockwood and Stamper (1999) have shown that the annual means of I from each are very well correlated with the annual means of F_s.

This point is demonstrated in figure 8. Using inter-calibration factors for the various instruments (e.g. Willson 1997), Lockwood and Stamper find a strong (correlation coefficient $c=0.852$) and highly significant $(100-6.15\times10^{-10}\%)$ relationship between I and F_s. If we assume that this is valid at all times, we can then use the best-fit regression

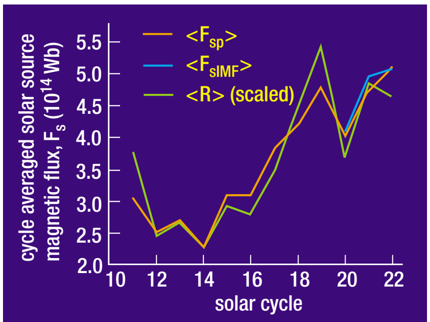
$$[I \text{ in W m}^{-2}] = 1364.9 + s\times[F_s \text{ in Wb}]\times10^{-14}$$

(where the slope $s=0.507\pm0.070$) to extrapolate back to 1868. Given than the heat capacity of the oceans will smooth out most of the effects of variations in I on the timescales of the solar cycle and shorter (Wigley and Raper 1990), we here look at 11-year running means, I_{11}; these smoothed variations being the most relevant to global temperature change. We find that the average total solar irradiance I_{11} increased by $\Delta I_{11}=1.65\pm0.23$ W m^{-2} in the interval 1901–1995, up to 1367.6 W m^{-2}. This is slightly lower than, but remarkably close to, the $\Delta I_{11}=2.106$ W m^{-2} estimated by Lean *et al.* (1995) using a solar cycle variation added to a long-term drift obtained by comparing the Sun's Maunder minimum to the luminosity of non-cyclic, sun-like stars.

The agreement between the forms of the two extrapolations is remarkably close, considering we have used an entirely independent set of measurements, namely the *aa* geomagnetic index. Furthermore, Lean and co-workers have recently refined their estimate and this has reduced their long-term drift factor by 25%, making their reconstruction of the variation of I almost identical to that shown here (J Lean, private communication 1999). The mean rise in I over the last three solar cycles is at a rate of 0.25 ± 0.4 W m^{-2} per decade. We can compare this range with the estimates made from inter-calibrated measurements during the minima at the start of cycles 22 and 23 by Willson (1995). He reported 0.50 and 0.37 W m^{-2} per decade for ACRIM1/2 and ERBS observations, respectively. Thus our estimates of the recent rise in I are comparable with (but somewhat smaller than) those by Willson. The reconstruction of solar brightness found here is also similar in its overall drift to that by Hoyt and Schatten (1993) and others, based on the length of the solar cycle. However, there are also some significant differences.

The $0.12\pm0.02\%$ rise in I since 1900 that is reported here is significant, being larger than the amplitude of the solar cycle variation and, because it is over a much longer timescale, its effects will not be smoothed out by factors such as the heat capacity of the oceans (Wigley and Raper 1990). It gives a change in the radiative forcing at the top of the atmosphere of $\Delta Q=\Delta I(1-a)/4\approx0.29\pm0.04$ W m^{-2}, where a, the Earth's albedo, is here taken to be 0.3 (see discussion of the concept of Q by Hansen *et al.* 1997). The effect of any change in I on global mean surface temperatures will be complex because it will be made up of contributions that are much stronger at some wavelengths (for example, UV) than at others and because a variety of other effects (for example, changes in anthropogenic greenhouse gases, tropospheric sulphate aerosols and volcanic dust in the stratosphere, ozone absorption of UV etc) will also be active and will interact with each other in complex feedback loops (Rind and Overpeck 1993; Hansen *et al.* 1997). This estimate of ΔQ due to solar change is similar to the estimate by the IPCC (Intergovernmental Panel on Climate Change). By way of comparison, the IPCC's ΔQ estimates for the same interval due to CO_2, other greenhouse gasses, and aerosols are roughly 1.5 W m^{-2}, 1.1 W m^{-2}, and –1.3 W m^{-2}, respectively (e.g. Wigley *et al.* 1997).

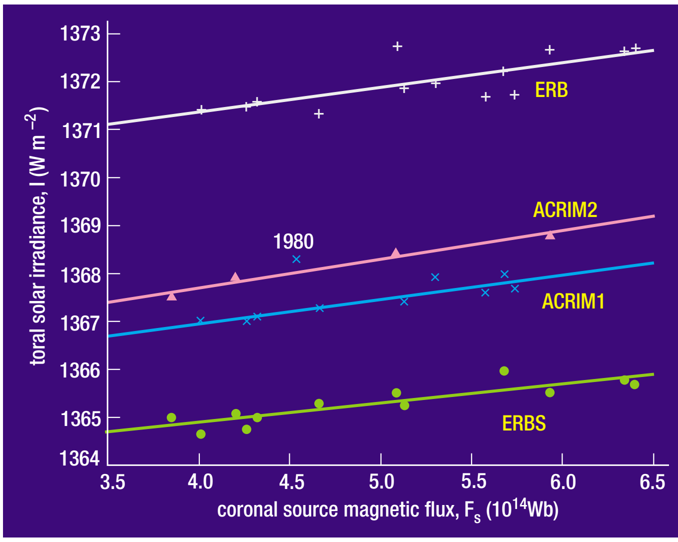
In order to evaluate the effect of the solar brightness change, we need to know how sensitive the Earth's climate system is to changes in the radiative power, Q. Figure 9 shows the



7: The solar cycle averages (minimum to minimum) of the solar source flux predicted from *aa*, <F_{sp}> (in orange) and from IMF observations <F_s> (in blue). Also shown are the mean sunspot numbers <*R*> (in green), which have been scaled using the least-squares regression <R> = 20.287{[<F_{sp}> in Wb]/10^{14}} − 15.238: the correlation coefficient is 0.927, which is significant at the (100 − 7.135 × 10^{-4})% level.

Table 2: Comparison of F_s observed from IMF and F_{sp} estimated using *aa* for solar cycles 20, 21 and 22

solar cycle no.	fitted or test data?	<F_s> (10^{14} Wb)	<F_{sp}> (10^{14} Wb)	(<F_s> − <F_{sp}>) (10^{14} Wb)	% error $\varepsilon = (100/\langle F_s\rangle) \times \lvert\langle F_s\rangle - \langle F_{sp}\rangle\rvert$	<$(F_s - F_{sp})^2$>$^{1/2}$ (10^{14} Wb)
20	test	4.0881	4.0253	0.0628	1.54	0.4930
21	fitted	4.9555	4.7316	0.2239	4.52	0.5527
22	fitted	5.0685	5.1087	−0.0402	0.79	0.4855



8: Scatter plots of the total solar irradiance, *I*, as a function of 3-year running means of annual coronal source magnetic flux values, F_s, for data from the instruments: Nimbus/ERB (+); SMM/ACRIM1 (×); UARS/ACRIM2 (Δ); and ERBS (o). The lines are linear regression fits to each data set (after Lockwood and Stamper 1999).

results for 11-year running means with various estimates of the "climate sensitivity", dT/dQ, where T is the global mean of the surface temperature. The orange line, bounded by uncertainty limits in green, is the inferred contribution of the increase in solar luminosity, on its own, ΔT_{i1} using the mean regression slope *s* of 0.507 ± 0.070 with a climate sensitivity estimate of dT_{i1}/dQ of 0.85 ± 0.15 °C/W m^{-2} (Rind and Overpeck 1993; Lean *et al.* 1995), giving

$$\Delta T_{i1} = -0.16 + (0.147 \pm 0.026) \times ([I \text{ in W m}^{-2}] - 1365)$$

This represents a consensus view of the Earth's climate sensitivity, being the average of values from several large numerical models of the coupled atmosphere–ocean circulation system, with an uncertainty set by the range of the estimates. From this we infer that the Sun's brightness change, on its own, could have caused a temperature rise of 0.24 ± 0.04 °C since 1900. The yellow line employs one of the largest published estimates of the climate sensitivity, $dT_{i2}/dQ = 1.7$ °C/W m^{-2}, by Nesme-Ribes *et al.* (1993). This gives

$$\Delta T_{i2} = -0.70 + 0.293 \times ([\text{I in W m}^{-2}] - 1365)$$

The inferred variations are very highly correlated with the observed global average of the surface temperature, ΔT_o (blue line) (Wigley *et al.* 1997), the correlation coefficient being 0.93 and at a lag of $\delta t = 2$ years, which is consistent with the heat storage effect of the oceans (Wigley and Raper 1990). The red line is a best linear regression fit to ΔT_o,

$$\Delta T_f = -0.48 + 0.385 \times ([I \text{ in W m}^{-2}] - 1365)$$

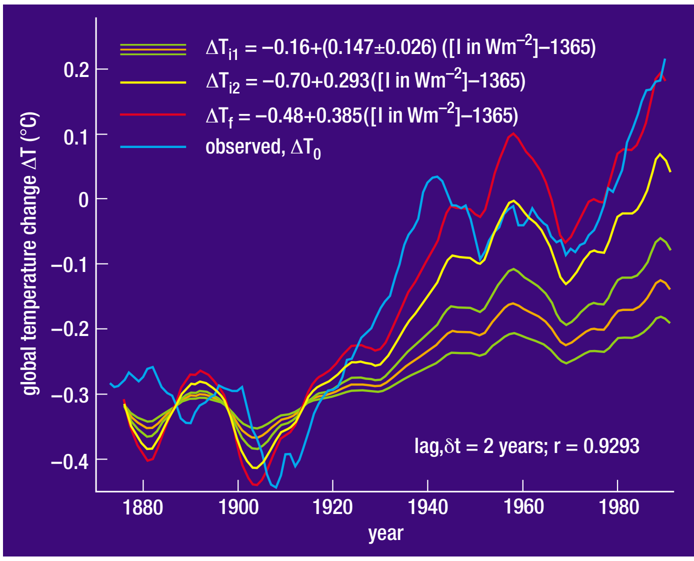
which corresponds to a climate sensitivity of $dT_f/dQ = 2.2$ °C/W m^{-2}, which would be required to explain the observed temperature rise ΔT_o in terms of solar irradiance variations alone. This is a higher value than any of the published estimates from modelling studies and roughly twice the consensus value.

Great care must be taken not to over-interpret this correlation. It undoubtedly argues for some solar variability effect on global warming, but adding an approximately exponential contribution (expected for man-made greenhouse gases and aerosols) does not decrease the correlation and can even make it higher (Laut and Gundermann 1998). Thus the high correlation does not support a purely solar effect. Multi-variable analysis that accounts for several mechanisms, including anthropogenic effects, does reveal improved correlations if some solar drift is included (Wigley *et al.* 1997; Tett *et al.* 1997; Benestad 1999).

If we compare the inferred temperature rise ΔT_{i1} (for the average prediction of the climate sensitivity) with that observed ΔT_o we find no significant difference for the period 1870–1910. On the other hand, the change in solar luminosity alone can account for only 52% of the rise in ΔT_o over the period 1910–1960, but just 31% of the rapid rise in ΔT_o over 1970–present. In the interval covered by figure 9, industrially-produced CO_2 in the atmosphere increased from about 280 to 355 ppmv. The implications are that the onset of a man-made contribution to global warming was disguised by the rise in solar brightness and that the anthropogenic effect may have a later, but steeper, onset than previously thought. Such an effect is consistent with the predictions for combined greenhouse and aerosol pollutants (Wigley *et al.* 1997; Hansen *et al.* 1997). Recently, Tett *et al.* (1999) have used a set of simulations made by a coupled atmosphere–ocean global circulation model to deduce a shift from solar forcing to anthropogenic effects as this century has progressed.

There is other evidence to support the view that solar changes have had a significant effect. For example, the decrease in global temperatures during about 1950–1965 was at a time when the concentration of greenhouse gasses was increasing (Friis-Christensen and Lassen 1991) and anthropogenic climate forcing was increasing (Wigley *et al.* 1997). The decrease in solar luminosity helps explain this and consequently including solar forcing can improve the correlation found with observed temperature (Wigley *et al.* 1997). Furthermore, the present upward trend in global temperatures appears to have commenced considerably earlier than significant burning of fossil fuels (Bradley and Jones 1993) and there is some evidence that temperatures have been as high in past epochs as they are now (see Cliver *et al.* 1998).

From the above, we can conclude that for solar brightness change to explain all of global warming would require a climate sensitivity considerably higher than the values predicted (by a factor of about two on average). However, there is a second mechanism whereby F_s would influence the global climate that has been proposed. This is more controversial and is based on the observation that total terrestrial cloud cover since 1984 is correlated with the flux of galactic cosmic rays reaching the Earth (Svensmark and Friis-Christensen 1997). Furthermore, this effect is strongest nearer the poles where cosmic rays have easier access to the atmosphere, the shielding by the Earth's magnetic field being less efficient there. The



9: Global temperature change, demonstrated by 11-year running averages of the global mean of the observed surface temperature, ΔT_o (blue line). The average of the regressions shown in figure 8 yield a total solar irradiance of [I in W m^{-2}] = 1364.9 + $s \times$ [F_s in Wb] $\times 10^{-14}$ and the inferred contributions assume that the radiative forcing at the top of the atmosphere is $Q = I(1-a)/4$ for an Earth albedo a of 0.3. The orange line is the inferred contribution of the increase in solar luminosity, ΔT_{i1} using the mean slope s of 0.507 with a climate sensitivity dT_{i1}/dQ of 0.85 °C/W m^{-2} as used by Lean *et al.* (1995) and the green lines deliniate the uncertainty caused by the uncertainty in s of ±0.070 and the likely range of dT_{i1}/dQ of ±0.15 °C/W m^{-2} (Rind and Overpeck 1993). The yellow line employs one of the largest published estimates of the climate sensitivity dT_{i2}/dQ = 1.7 °C/W m^{-2}, by Nesme-Ribes *et al.* (1993). The red line is a best linear regression fit to ΔT_o, ΔT_f, and shows that an even higher climate sensitivity of $dT_f/dQ = 2.2$ °C/W m^{-2} is required to explain the observed temperature rise ΔT_o in terms of solar irradiance variations alone. The peak correlation is 0.93, at a lag of $\partial t = 2$ years. All ΔT predictions are relative to the mean ΔT_o for solar cycle 13 (1889–1901).

implication is that cosmic rays may be involved in the production of at least some clouds. The flux of galactic cosmic rays is well known to be anticorrelated with the sunspot numbers because the Earth is more shielded from cosmic rays at sunspot maximum by the heliosphere. More than one mechanism contributes to this shielding: the particles are scattered by irregularities in the heliosphere; they are convected and decelerated by the solar wind; and they suffer gradient and curvature drifts because of the large-scale structure of the heliospheric field.

The relative importance of the various effects is not yet known (see review by Moraal 1993) although Cane *et al.* (1999) have shown that the cosmic-ray fluxes anticorrelate well with the coronal source flux. Stamper *et al.* (1999) have reported that the fluxes do show a significant long-term drift to lower values consistent, at least qualitatively, with the rise in heliospheric field which by equations (3) and (4) is proportional to F_s; however, the percentage long-term change in cosmic-ray flux is only about a tenth of that in F_s. From the 140% increase in mean F_s over this century, we would predict a drop in cosmic-ray fluxes of 17%, comparable with the amplitude of the variation over solar cycle 21. Svensmark and Friis-Christensen (1997) report a variation in total cloud cover between about −1% and +2% over this cycle. From the above, we would expect this mechanism, were it to be effective, to give roughly a 3% decrease in average cloud cover since 1900. The net effect on the terrestrial climate is unclear as there are two competing effects: increased cloud cover increases the Earth's albedo, and so reflects more solar power back into space, but also traps heat.

The past and the future

The method reported here allows us to study the changes in the solar magnetic field and brightness since the middle of the last century when regular geomagnetic monitoring began. Longer-term variations in the Sun have been inferred from historical sunspot and auroral observations (Hoyt and Schatten 1992 and Silverman 1992) and from the terrestrial abundances of isotopes such as ^{14}C and ^{10}Be (produced by cosmic-ray bombardment and deposited and stored, for example, in the polar icecaps) (Sonnet 1991, Beer *et al.* 1998). The cosmogenic ^{10}Be data, in particular, show a century-long decay in cosmic-ray fluxes, with a variation somewhat similar to that in the coronal source field shown in figure 6 (Lean *et al.* 1995). The isotope data also show that solar activity can largely disappear for periods of 50–100 years, such as the Maunder minimum (circa 1650–1700), although there is evidence that a weak and cyclic magnetic field still emerged from the Sun during the minimum itself (Beer *et al.* 1998). By comparing the phase of the 88-year "Gleissberg" solar oscillation prior to and after the Maunder minimum, it has been inferred that the dynamo generating the solar field may be chaotic rather than quasi-periodic (Feynman and Gabriel 1990). Such chaotic behaviour may be the cause of the sudden changes in F_s around 1900 and 1960 that are evident in figure 6.

Recently, the SOHO satellite has provided us with a veritable wealth of new information concerning the Sun (Priest 1998). The challenge now is to use that information to understand the origin of the long-term solar changes reported here – to the point where we can predict them, their effects on our operational systems and their impact on our climate. ●

M Lockwood, R Stamper, M N Wild (World Data Centre C-1 for STP, Rutherford Appleton Laboratory, Chilton, UK). A Balogh and G Jones (Blackett Laboratory, Imperial College, London, UK). *This research is supported by PPARC. The authors are also grateful for valuable discussions with J Haig, M Allen and J Lean.*

References

Baker D 1986 in *Solar Wind–Magnetosphere Coupling* edited by Y Kamide and J A Slavin, 17–38, Terra Scientific, Tokyo.
Balogh A *et al.* 1995 *Science* **268** 1007–1010.
Beer J *et al.* 1998 *Solar Physics* **181** 237–249.
Benestad R A 1999 *A&G* **40** 3.14–3.17.
Blanford H F 1891 *Nature* **43** 583–587.
Bradley R S and P D Jones 1993 *The Holocene* **3(4)** 367–376.
Cane H V *et al.* 1999 *Geophys. Res. Lett.* **26** 565–568.
Clilverd M A *et al.* 1998 *J. Atmos. Sol.-Terr. Phys.* **60** 1047–1056.
Cliver E W *et al.* 1996 *J. Geophys. Res.* **101** 27 091–27 109.
Crooker N U and K I Gringauz 1993 *J. Geophys. Res. 98* 59–62.
Feynman J and N U Crooker 1978 *Nature* **275** 626–627.
Feynman J and S B Gabriel 1990 *Solar Physics* **127** 393–403.
Friis-Christensen E and K Lassen 1991 *Science* **245** 698–700.
Gazis P R 1996 *Rev. Geophys.* **34** 379–402.
Gringauz K I 1981 in Solar Wind 4, *Report MPAE-W-100-81*-31 edited by H Rosenbauer, MPI für Aeronomie, Lindau, Germany.
Hansen J *et al.* 1997 *J. Geophys. Res.* **102** 6831–6864.
Hapgood M A 1993 *Ann Geophys.* **11** 248.
Herschel W 1801 *Phil. Trans. R. Soc. (London)* **91** 265–318 and 354–362.
Hoyt D V and K H Schatten 1993 *J. Geophys. Res.* **98** 18 895–18 906.
Laut P and J Gundermann 1998 *J. atmos. sol.-terr. Phys.* **60** 1–3.
Lean J *et al.* 1995 *Geophys. Res. Lett.* **22** 3195–3198.
Lockwood M 1997 *A&G* **38(1)** 21–25
Lockwood M and R Stamper 1999 *Geophys Res. Lett.* in press.
Lockwood *et al.* 1999a *Nature* **399** 437–439.
Lockwood M *et al.* 1999b *J. Navigation* **52** 203–216.
Mayaud P N 1972 *J. Geophys. Res.* **72** 6870–6874.
Moraal H 1993 *Nuclear Phys. B (Proc. Suppl.)* **33A,B** 161–178.
Nesme-Ribes E *et al.* 1993 *J. Geophys. Res.* **98** 18 923–18 935.
Priest E 1998 *A&G* **39** 3.10–3.13.
Pulkkinen P J *et al.* 1999 *Astron. Astrophys.* **341** L43–L46.
Russell C T 1975 *Solar Physics* **42** 259–269.
Rind D and J Overpeck 1993 *Quaternary Sci. Rev.* **12** 357–374.
Sargent H H III 1986 in *Solar Wind–Magnetosphere Coupling* edited by Y Kamide and J A Slavin, 143–148, Terra Scientific, Tokyo.
Schimel D *et al.* 1996 in *Climate Change, 1995: The Science of Climate Change* Chapter 2, Cambridge University Press.
Scurry L and C T Russell 1991 *J. Geophys. Res.* **96** 9541–9548.
Silverman S W 1992 *Rev. Geophys.* **30** 333–351.
Sonnet C P 1991 in U.S. National Report to IUGG, 1987–1990, Suppl. to *Rev. Geophys.* 909–914.
Stamper *et al.* 1999 *J. Geophys Res.* in press
Svensmark H and E Friis–Christensen 1997 *J. atmos. Sol. Terr. Phys.* **59** 1225–1232.
Tett S *et al.* 1999 *Nature* in press.
Vasyliunas V M *et al.* 1982 *Planet Space Sci.* **30** 359–365.
Wang Y-M *et al.* 1996 *Science* **271** 464–469.
Wang Y-M and N R Sheeley, Jr 1995 *Astrophys. J.* **447** L143–L146.
Webb D F and R A Howard 1994 *J. Geophys. Res.* **99** 4201.
Wigley T M L and S C B Raper 1990 *Geophys. Res. Lett.* **17** 2169–2172.
Wigley T M L *et al.* 1997 *Proc. Natl. Acad. Sci., USA* **94** 8314–8320.
Willson R C 1997 *Science* **277** 1963–1965.