II. Full Solar Cycle Simulations

D. H. MACKAY\textsuperscript{1} and M. LOCKWOOD\textsuperscript{2}

\textsuperscript{1} School of Mathematics and Statistics, University of St. Andrews, St. Andrews, Fife, Scotland, KY16 9SS
\textsuperscript{2} World Data Centre C1, Rutherford Appleton Laboratory, Chilton, Oxfordshire, UK

(Received 18 April 2002; accepted 20 June 2002)

Abstract. In this paper the origin and evolution of the Sun’s open magnetic flux is considered by conducting magnetic flux transport simulations over many solar cycles. The simulations include the effects of differential rotation, meridional flow and supergranular diffusion on the radial magnetic field at the surface of the Sun as new magnetic bipoles emerge and are transported poleward. In each cycle the emergence of roughly 2100 bipoles is considered. The net open flux produced by the surface distribution is calculated by constructing potential coronal fields with a source surface from the surface distribution at regular intervals. In the simulations the net open magnetic flux closely follows the total dipole component at the source surface and evolves independently from the surface flux. The behaviour of the open flux is highly dependent on meridional flow and many observed features are reproduced by the model. However, when meridional flow is present at observed values the maximum value of the open flux occurs at cycle minimum when the polar caps it helps produce are the strongest. This is inconsistent with observations by Lockwood, Stamper and Wild (1999) and Wang, Sheeley, and Lean (2000) who find the open flux peaking 1–2 years after cycle maximum. Only in unrealistic simulations where meridional flow is much smaller than diffusion does a maximum in open flux consistent with observations occur. It is therefore deduced that there is no realistic parameter range of the flux transport variables that can produce the correct magnitude variation in open flux under the present approximations. As a result the present standard model does not contain the correct physics to describe the evolution of the Sun’s open magnetic flux over an entire solar cycle. Future possible improvements in modeling are suggested.

1. Introduction

The Sun’s open magnetic flux which originates in coronal holes (Wang and Sheeley, 1990) extends outwards from the Sun into interplanetary space. In interplanetary space it forms the interplanetary magnetic field (IMF) which surrounds the Earth and interacts with the Earth’s magnetosphere. In recent years there has been much interest in studying through observations and theory both the open flux and the IMF since variations in both have been linked to variations in the Earth’s climate (Svensmark and Friis-Christensen, 1997; Svensmark, 1998; Bond \textit{et al}., 2001; Lockwood 2001, 2002) as well as to many effects in the near-Earth plasma environment studied as part of ‘space weather’. To understand fully and predict how the open magnetic flux and IMF may affect us we need to understand its.
origin and variation as the Sun’s surface magnetic flux varies through its 11-year activity cycle.

Several papers in recent years have aimed to determine how the open flux and near-Earth IMF vary over both the short term (few years) and also the long term (many cycles). Variations in the total open flux and the near-Earth IMF can be considered as equivalent because the Ulysses spacecraft has shown that the radial field in the heliosphere is close to being independent of heliographic latitude (Balogh et al., 1995). Lockwood et al. (2002) have used the Ulysses data to show that the ratio of the total open solar flux to the radial IMF component is constant to within 5%, at both sunspot maximum and minimum. Lockwood, Stamper, and Wild (1999) and Wang, Lean, and Sheeley (2000) showed that the open magnetic flux (and therefore the IMF) varies throughout the solar cycle. They found that with each solar cycle the modulation of the open flux lags the total surface flux (and sunspot number) by 1–2 years (see also Wang and Sheeley, 2002). Lockwood, Stamper and Wild (1999) also found that the maximum magnetic flux leaving the Sun varies strongly from one cycle to the next with an increase in average values of 2.3 since 1901. Wang, Lean, and Sheeley (2000) determined their variation through the reconstruction of potential magnetic fields from observed synoptic magnetograms, while Lockwood, Stamper, and Wild (1999) used the indirect method of the aa index (Mayaud, 1971). To determine why the open flux lags behind the surface flux, Wang, Sheeley, and Lean (2000) developed a magnetic flux transport model (Wang, Nash, and Sheeley, 1989a; van Ballegooijen, Cartledge, and Priest, 1998) where the radial magnetic field at the solar surface is evolved under the effects of differential rotation, meridional flow and supergranular diffusion for a number of years. With this they showed for one or two large bipoles how the open flux may evolve independently of the surface flux due to the effect of differential rotation and gave a suggestion why the observed lag should occur. Mackay, Priest, and Lockwood (2002) extended this by considering in detail how the tilt angle and latitude of emergence of a single bipole could affect the amount and variation of the surface and open flux. In contrast, Solanki, Schüssler, and Fligge (2000) constructed a semi-empirical model which relates the open flux emergence rate to the observed sunspot number and assumed a linear loss rate of open flux with a best-fit time constant. With this approach a good representation of how the open flux has varied over the last 300 years could be found.

In the paper by Wang, Sheeley, and Lean (2000) it was hypothesised that the observed lag between the surface and open flux is a result of the fact that at sunspot maximum activity is distributed more or less uniformly over longitude and magnetic polarities are mixed. These closely mixed polarities limit by strong cross connections the amount of open magnetic flux. However, as flux is canceled, activity becomes concentrated in one or two major complexes and longitudinal polarity separations increase. This increase in separation can then lead to an increase in open flux 1–2 years after sunspot maximum.
In this paper we continue the work of Wang, Sheeley, and Lean (2000) and Mackay, Priest, and Lockwood (2002) to consider in detail how the magnitude of the open flux of the Sun varies as multiple bipoles emerge, interact and are transported poleward over many solar cycles. The production and location of coronal holes under similar circumstances can be seen in Wang and Sheeley (1990). The paper is outlined as follows. In Section 2 the model is described. The input data used to simulate the solar cycles is discussed in Section 3. In Section 4 the simulations are carried out for seven, 11-year solar cycles and the variation of the open flux is calculated as key parameters are varied. Finally, in Section 5 the results and consequences for future modeling are discussed.

2. The Model

To consider the evolution of the Sun’s open magnetic flux we shall use a magnetic flux transport model (DeVore, Sheeley, and Boris, 1984; Sheeley, DeVore, and Boris, 1985; Sheeley, Nash, and Wang, 1987; Wang, Nash, and Sheeley 1989a; van Ballegooijen, Cartledge, and Priest, 1998; Mackay, Priest, and Lockwood, 2002). The model evolves the radial component of the Sun’s magnetic field at the solar surface under the combined effects of flux emergence, differential rotation, meridional flow and supergranular diffusion. Let \( B_r (R_\odot, \theta, \phi, t) \) be the radial magnetic field at \( r = R_\odot \), where \( r \) is the radial distance from the Sun’s center, \( \theta \) the polar angle, \( \phi \) the azimuthal angle and \( t \) time. Here \( B_r \) represents the large-scale field of the Sun which is averaged over spatial scales larger than a supergranule (30 Mm). The evolution of the field at the solar surface, \( (r = R_\odot) \), is then described by the equation

\[
\frac{\partial B_r}{\partial t} = \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left[ \sin \theta \left( -u(\theta) B_r + D \frac{\partial B_r}{\partial \theta} \right) \right] + \frac{D}{\sin^2 \theta} \frac{\partial^2 B_r}{\partial \phi^2} - \frac{\Omega(\theta)}{\Omega_1(\theta)} \frac{\partial B_r}{\partial \phi},
\]

(1)

where \( u(\theta) \) is the meridional flow, \( \Omega(\theta) \) the differential rotation profile and \( D = 600 \text{ km s}^{-1} \) the photospheric diffusion constant (Leighton, 1964).

The meridional flow which is directed poleward is given as a function of latitude \( (\lambda = \pi/2 - \theta) \) by

\[
u(\lambda) = \begin{cases} 
- u_0 \sin (\pi \lambda/\lambda_0) & | \lambda | < \lambda_0, \\
0 & \text{otherwise},
\end{cases}
\]

(2)

so above \( \lambda_0 \) the flow velocity vanishes. The particular values for these constants are \( \lambda_0 = 75^\circ \) and \( u_0 = 11 \text{ m s}^{-1} \) (Hathaway, 1996; Snodgrass and Dailey, 1996). The differential rotation profile is given by (Snodgrass, 1983)

\[
\Omega(\theta) = 13.38 - 2.30 \cos^2 \theta - 1.62 \cos^4 \theta - 13.20 \text{ deg day}^{-1}.
\]

(3)

For simplicity and to provide information on the evolution of the dipole components of the field (Wang, Sheeley, and Lean, 2000; Mackay, Priest, and Lockwood,
Figure 1. Input data for the solar cycle simulations. (a) The butterfly diagram. (b) Rate of flux emergence per 27 days. (c) Relative numbers of positive tilts (solid line with crosses), zero tilts (solid line) and negative tilts (dashed line). (d) Average tilt angle over all cycles with respect to latitude of emergence.
2002; Wang and Sheeley, 2002), the radial magnetic field is expressed in terms of spherical harmonic functions as

\[ B_r(r, \theta, \phi, t) = \sum_{l=1}^{N} b_l(r, \theta, \phi, t) = \sum_{l=1}^{N} \sum_{m=0}^{l} b_{lm}(r, \theta, \phi, t), \]

where \( b_l(r, \theta, \phi, t) \) represents a multipole, and \( b_{lm}(r, \theta, \phi, t) \) represents each spherical harmonic component at the radius \( r \), where \( l \) is the harmonic degree and \( m \) is the azimuthal mode number. For the simulation \( N = 63 \), which is sufficient to resolve structures on the size of a supergranule (30 Mm).

As the surface field is evolved, a coronal magnetic field is extrapolated from it at regular 27-day intervals to determine the amount of open flux. At each 27-day interval the instantaneous surface map which represents the simulated magnetic field of the entire Sun at that instant is used. Therefore this study differs from that of Wang, Lean, and Sheeley (2000) since they used synoptic data which does not take into account temporal variations between the initial and final synoptic longitudes which lie 27.3 days apart. The coronal field obtained from \( B_r(R_\odot, \theta, \phi) \) is assumed to be potential (\( \nabla \times B = 0 \)) with a source surface at \( r = R_{ss} = 2.5 R_\odot \) where \( B_\theta = B_\phi = 0 \). At the source surface the field is assumed to become purely radial and field lines extending out to it are classified as open. The source surface which is widely used crudely simulates the effect of the solar wind opening up magnetic field lines and producing coronal holes (Wang and Sheeley, 1990). With this approximation the relationship between harmonics at different radii (van Ballegooijen, Cartledge, and Priest, 1998) is

\[ b_{lm}(r, \theta, \phi) = b_{lm}(R_\odot, \theta, \phi) \left[ \frac{(l+1)(r/R_\odot)^{-l-2}}{l+1+l(R_{ss}/R_\odot)^{-2l-1}} + \frac{l(R_{ss}/R_\odot)^{-2l-1}(r/R_\odot)^{l-1}}{l+1+l(R_{ss}/R_\odot)^{-2l-1}} \right]. \]

The potential approximation fixes the radial dependence of the harmonics and it is clear that the higher the \( l \) value the faster the multipole, \( l \), falls off. Therefore, only the lower order multipoles have a significant contribution to the field which extends out to the source surface. In calculating the open flux it is important to distinguish between the flux transport model used to produce the surface harmonics \( (b_{lm}(R_\odot, \theta, \phi)) \) and the potential source surface model which then extrapolates these harmonics to higher heights \( (b_{lm}(r, \theta, \phi)) \). Both models complement one another but are clearly independent of each other.

As with Wang, Sheeley, and Lean (2000) and Mackay, Priest, and Lockwood (2002), we shall determine the following variables: the total (net) surface flux (Equation (6)), the total (net) open flux (Equation (7)), the non-axisymmetric dipole component (Equation (8)) at \( R_{ss} \) and the total dipole component (Equation (9)) at \( R_{ss} \).
\[
\Phi_{\text{tot}} = R_\odot^2 \int |B_r(R_\odot, \theta, \phi, t)| \, d\Omega, \tag{6}
\]

\[
\Phi_{\text{open}} = R_{ss}^2 \int |B_r(R_{ss}, \theta, \phi, t)| \, d\Omega, \tag{7}
\]

\[
\langle b_{11}(R_{ss}) \rangle = \frac{\int |b_{11}(R_{ss}, \theta, \phi, t)| \, d\Omega}{4\pi}, \tag{8}
\]

\[
\langle b_1(R_{ss}) \rangle = \frac{\int |b_1(R_{ss}, \theta, \phi, t)| \, d\Omega}{4\pi}. \tag{9}
\]

The non-axisymmetric dipole represents the lowest order east–west dipole component at the source surface, while the total dipole is the sum of the lowest order north–south \((b_{10}, \text{axisymmetric dipole})\) and east–west \((b_{11}, \text{non-axisymmetric dipole})\) component at the source surface. In the next section the input data used to simulate the solar cycles is discussed.

### 3. Input Data and Initial Flux Distribution

When new magnetic flux emerges on the Sun it does so with a wide range of sizes and strengths (Wang and Sheeley, 1989; Harvey and Zwaan, 1993). To consider the evolution of the open flux over many solar cycles we are only going to consider large-scale emergence of flux (Gaizauskas et al., 1983) and neglect the small-scale magnetic carpet. These large-scale emergences follow an approximately eleven-year cycle and interact with one another to produce large unipolar areas which are transported across the solar surface. The smaller flux elements (such as ephemeral regions; Harvey, 1984) with random orientations are not included, since they are likely to connect very low down in the solar atmosphere and are unlikely to contribute significantly to the open flux which reaches out to \(R_{ss} = 2.5 R_\odot\).

#### 3.1. Statistical Input Data

To simulate the evolution of the Sun’s surface and open magnetic flux over many solar cycles realistic input data on the statistical properties of new emerging bipolar magnetic regions is required. This is important not only for obtaining the correct surface pattern but also for the surface fluxes relation to the open flux (Equation (5)). In Mackay, Priest, and Lockwood (2002) it was shown in detail how the amount and variation of the open magnetic flux of a single bipole depends on its latitude of emergence and initial tilt angle \(\alpha\). The tilt angle is the angle between the line joining the centers of the polarities and the east–west line. Positive tilt angles denote a leader flux lying equatorward of the follower flux, in agreement with Joy’s...
If realistic simulations are to be carried out accurate input data on the latitude of emergence, relative number of positive versus negative tilt-angle bipoles and variation of tilt-angle with latitude of emergence is required. Such information can be found in the papers by Wang and Sheeley (1989) and Tian et al. (1999).

In Figure 1 properties of the input data for new bipolar magnetic regions can be seen where the simulation is followed for seven, 11-year solar cycles. In Figure 1(a) the butterfly diagram is shown. At the start of each cycle new flux emerges at roughly 40° latitude in each hemisphere. As the cycle progresses the emergence latitude decreases and approaches 5° by the end. The width of each wing of the butterfly diagram is 10° at the start of a cycle and decreases to roughly 5° by the end. The assumed cycle period is 11 years with the length of each cycle 13 years. This gives an overlap of 1 year between each old and new cycle (van Ballegooijen, Cartledge, and Priest, 1998). As with van Ballegooijen, Cartledge, and Priest (1998) the simulation starts at the end of a cycle just when the first regions of a new cycle start to emerge. In each cycle roughly 2100 new bipolar magnetic regions are assumed to emerge at random longitudes. In Figure 1(b) the rate of flux emergence taken over a 27-day period is shown. This closely resembles the observed sunspot number, at cycle maximum there is just over 1 region emerging per day which drops off to roughly 1 every 15 days at cycle minimum. The number of bipoles emerging each year with positive (α ≥ 0.5°), zero (−0.5° < α < 0.5°) and negative (α ≤ −0.5°) tilt angles is given in Figure 1(c). The solid line with/without the crosses denotes the positive/zero tilt-angle cases and the dashed line the negative tilt-angle case. The relative numbers of each match well those seen in a similar plot by Wang and Sheeley (1989). In close agreement with Wang and Sheeley (1989), 78% of the simulated bipoles have tilt angles in the range −10 to +30°. The average tilt angle varies as 0.5λ, where λ is the latitude of emergence, which also fits the results determined by Wang and Sheeley (1989). The root mean square value of α is 18°. From this it can be seen that the input data is consistent with observations.

### 3.2. DESCRIPTION OF BIPOLES

When each new bipolar magnetic region emerges at random longitude it adds a contribution \( \delta B_r(R_\odot, \theta, \phi) \) to the radial magnetic field at the solar surface (see van Ballegooijen, Cartledge, and Priest, 1998),

\[
\delta B_r(R_\odot, \theta, \phi) = B_r^+(R_\odot, \theta, \phi) - B_r^-(R_\odot, \theta, \phi),
\]

where

\[
B_r^\pm(R_\odot, \theta, \phi) = B_{\text{max}}e^{-2[1-\cos\beta_{\pm}(\theta, \phi)]/\rho_0^2}.
\]

If the central positions of the positive and negative poles of the new bipole are given by \((\theta_+, \phi_+)\) and \((\theta_-, \phi_-)\), then \(\beta_{\pm}(\theta, \phi)\) is defined as the heliocentric angle...
TABLE I

Parameters for new bipoles.

<table>
<thead>
<tr>
<th>Region</th>
<th>( \beta ) (deg)</th>
<th>( B_{\text{max}} ) (G)</th>
<th>Total (net) flux ( 10^{21} \text{Mx} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.0</td>
<td>250</td>
<td>17.1</td>
</tr>
<tr>
<td>2</td>
<td>7.8</td>
<td>152</td>
<td>9.3</td>
</tr>
<tr>
<td>3</td>
<td>6.1</td>
<td>92</td>
<td>4.9</td>
</tr>
<tr>
<td>4</td>
<td>4.7</td>
<td>56</td>
<td>2.5</td>
</tr>
<tr>
<td>5</td>
<td>3.7</td>
<td>34</td>
<td>1.2</td>
</tr>
</tbody>
</table>

between the point \((\theta, \phi)\) and \((\theta_\pm, \phi_\pm)\). For small \( \beta_\pm(\theta, \phi) \) the flux distribution is approximately Gaussian with \( \beta_0 \) the width of the Gaussian. In total, five separate sizes of bipoles are considered (van Ballegooijen, Cartledge, and Priest, 1998). Table I gives details on the separation in heliographic degrees of the centers of positive and negative flux \((\beta)\), the initial peak magnetic field strength \((B_{\text{max}})\) and the total (net) flux when the width of the Gaussian is chosen to be \( \beta_0 = 4^\circ \). The width is of sufficient size that a new bipole can be well resolved with the number of harmonic degrees used \((N = 63)\). In accordance with Harvey and Zwaan (1993) and Schrijver and Harvey (1994), the number of new bipolar magnetic regions emerging per unit time is assumed to be inversely proportional to the square of the region size such that the emergence rate is \( a(t)A^{-2} \, \text{d}A \), where \( A \) is the area of the new bipole (in square degrees) and \( a(t) \) is a function of time in the cycle. In each cycle a net flux of around \( 1 \times 10^{25} \text{Mx} \) emerges.

3.3. INITIAL FLUX DISTRIBUTION

To run the simulations an initial flux distribution of the radial magnetic field at the solar surface is required. As described by van Ballegooijen, Cartledge, and Priest (1998), the initial distribution is chosen to be one where there is a balance between equatorial diffusion and poleward meridional flow. The form of the radial magnetic field at the surface is independent of longitude and is given by

\[
B_r(R_\odot, \lambda) = \begin{cases} 
\text{sgn}(\lambda) B_0 e^{-a_0 \left[ \cos(\pi \lambda/\lambda_0) + 1 \right]} & | \lambda | < \lambda_0, \\
\text{sgn}(\lambda) B_0 & \text{otherwise},
\end{cases}
\]

(12)

where \( a_0 = (u_0 R_\odot \lambda_0)/(\pi D) \), \( \text{sgn}(\lambda) \) is the sign of the latitude and \( B_0 \) is the initial polar field strength. For the simulations here \( B_0 = 14.5 \text{G} \). This value is determined by the requirement that the peak value of the magnetic field in successive cycles remains roughly constant for the \( B_{\text{max}} \) values given in Table I and the emergence properties in Figure 1. This value is consistent with previously quoted values in the literature (Wang, Nash, and Sheeley, 1989b – 16 G; Wang and Sheeley,
Figure 2. Initial flux distributions for the full solar cycle simulations. (a) An axisymmetric distribution with balance between meridional flow and equator-ward diffusion. The pole strength is 14.5 G. (b) An initial distribution produced by the flux transport code after 67 years. In both cases white represents positive flux and black negative and the field saturates at 10 G.

1995 – 12 and 16 G). This flux distribution can be seen in Figure 2(a) where the top-knot profile commonly observed at solar minimum is present (white represents positive flux and black negative flux). As well as considering the axi-symmetric flux distribution another initial flux distribution is generated by running the flux transport simulation for 67 years from Figure 2(a) to remove its initial artificial nature. The properties of the new emerging bipoles are similar to those of Figure 1, but a different seed value for producing the random longitudes of emergence is used (compared to the simulations in Section 4). The distribution can be seen in Figure 2(b). The main differences between Figure 2(b) and Figure 2(a) are that the top-knot profile is now weaker and flux can be seen to be lying at lower latitudes.
4. Full Solar Cycle Simulations

The full solar cycle simulations are now carried out to determine the phase relationship between the open and surface flux as the flux transport model evolves the surface field and a potential model determines the amount of open flux. To begin with, all three surface effects of differential rotation, meridional flow and supergranular diffusion are included and then the role played by meridional flow (Wang, Nash, and Sheeley, 1989b) is considered by neglecting it.

4.1. Differential Rotation, Supergranular Diffusion and Meridional Flow

In Figure 3 the results of the full solar cycle simulation can be seen when the initial configuration is that of Figure 2(a). In each graph the quantities are determined once every 27 days. Figure 3(a) gives the total (net) surface flux ($\Phi_{\text{net}}/4\pi R_\odot^2$, solid line) and the total (net) open flux ($\Phi_{\text{open}}/4\pi R_\odot^2$, dashed line) throughout the seven solar cycles, where the total surface flux is given as an average value over the solar surface (in gauss, G) and the total open flux has been scaled to its equivalent strength (in nanotesla, nT) at the radius of the Earth ($R_e = 1$ AU). In expressing the open flux by its equivalent strength at the radius of the Earth (Wang, Sheeley, and Lean 2000), it is assumed that beyond the source surface the open flux is distributed uniformly in heliographic latitude and longitude. This assumption is supported by observations from Ulysses (Balogh et al., 1995; Lockwood et al., 1999; Smith et al., 2001). From the graph it can be seen that the variation of the two quantities is out of phase. The surface flux is in phase with the emergence rate (Figure 1(b)), while the open flux is out of phase and has maxima at cycle minima. The total open flux therefore behaves independently of the total surface flux. Since the peak in open flux occurs at cycle minimum the simulation is not consistent with the observations of Lockwood, Stamper, and Wild (1999) or Wang, Lean, and Sheeley (2000).

In Figure 3(b) the variation of the dipole components of the field at the source surface is given, where the solid line is the total dipole ($\langle b_1 \rangle(R_{ss}/R_e)^2$) and the dashed line represents the east–west non-axisymmetric dipole component ($\langle b_{11} \rangle(R_{ss}/R_e)^2$). As with the open flux, both have been scaled to their equivalent strengths (in nanotesla (nT)) at the radius of the Earth ($R_e = 1$ AU). Throughout each cycle the north–south axisymmetric component of the total dipole dominates over the east–west non-axisymmetric dipole component except for a short time around cycle maximum. As found by Wang, Sheeley, and Lean (2000) and Mackay, Priest, and Lockwood (2002) the total open flux follows the total dipole (in both phase and magnitude). Finally, in Figure 3(c) the variation of the polar magnetic fields in the northern hemisphere (solid line) and southern hemisphere (dotted line) is given. In each case the average magnetic field strength in gauss (G) is given for latitudes above 80° in each hemisphere. As expected, the polar fields reverse with
Figure 3. Results of the full solar cycle simulations when all three surface effects are included. (a) shows the evolution of the total surface flux ($\Phi_{\text{tot}}/4\pi R_E^2$ in gauss, solid line) and total open flux ($\Phi_{\text{open}}/4\pi R_E^2$ in nanotesla (nT), dashed line). (b) gives the evolution of the total dipole ($\langle b_1 \rangle (R_{ss}/R_E)^2$, solid line) and east–west non-axisymmetric dipole ($\langle b_{11} \rangle (R_{ss}/R_E)^2$, dashed line) in nanotesla. (c) Evolution of the average north polar field (solid line) and average south polar field (dotted line) in gauss for latitudes above 80°.
each 11-year cycle, as trailing polarity flux pushed poleward by meridional flow cancels the polar fields (Wang, Nash, and Sheeley, 1989b; and Wang and Sheeley, 1991). In each cycle the polar magnetic fields reverse 1–2 years after maximum determined from the bipole emergence rate given in Figure 1(b).

Both the total open flux and total dipole show a significant variation throughout a solar cycle. The maximum values (∼3 nT) occur at solar minimum when the polar field regions are the strongest. Their minimum values (∼1 nT) occur at solar maximum when the polar field regions are weak. This variation can be explained (Equation (5)) in terms of the surface flux distribution and the construction of the potential magnetic field, which relates harmonics of the surface magnetic field to those at higher heights (Section 4.3).

It is clear that when all three of the surface effects of differential rotation, meridional flow and supergranular diffusion are included at observed values, the simulation fails to produce the desired result of an open magnetic flux which peaks roughly 1–2 years after cycle maximum. We have found that the only way to make these simulations match the observed phase variation of the open flux is to effectively switch off meridional flow. We discuss the effects, implications and validity of this in the next section.

4.2. Differential Rotation and Supergranular Diffusion

We now consider the effect meridional flow has on the phase relationship between the surface and open magnetic flux. Meridional flow points from the equator to the pole in each hemisphere. In the simulation it is very weak, with a peak value of only 11 m s\(^{-1}\), but over the duration of a solar cycle it has an important role of pushing flux poleward (Wang, Nash, and Sheeley, 1989b) and aiding the cancellation and reversal of the poles. When switched off, magnetic flux may only extend poleward through diffusion. We now consider how the phase relationship between the surface and open magnetic flux (under the potential model assumption) varies when there is no meridional flow. As with Section 4.1, the initial flux distribution is chosen to be that of Figure 1(a) and the same input data is used. To consider how the north–south axisymmetric dipole strength at the solar surface of a single magnetic bipole behaves as the meridional flow velocity varies see Wang and Sheeley (1991).

In Figure 4, the results of the simulation can be seen where the same quantities are shown as in Figure 3. As previously found, the total surface flux (Figure 4(a)) varies in phase with the emergence rate and has a very similar behaviour to before. Even though it has a similar behaviour, the magnitude of the open flux at solar maximum and minimum is slightly less. In the absence of meridional flow the phase relationship between the open and surface flux has now changed. The open flux still has an 11-year period, but the maximum value now occurs between solar maximum and minimum. This is roughly the correct time to fit the observations of Lockwood, Stamper, and Wild (1999) and Wang, Lean, and Sheeley (2000). From Figure 4(b) it is clear that the open flux again closely follows the total dipole at the source
Figure 4. Results of the full solar cycle simulations when the meridional flow velocity is switched off. (a) Shows the evolution of the total surface flux ($\Phi_{\text{tot}}/4\pi R_\odot^2$ in G, solid line) and total open flux ($\Phi_{\text{open}}/4\pi R_\odot^2$ in nT, dashed line). (b) Gives the evolution of the total dipole ($\langle b_1 \rangle (R_{\odot}/R_e)^2$, solid line) and east-west non-axisymmetric dipole ($\langle b_{11} \rangle (R_{\odot}/R_e)^2$, dashed line) in nT. (c) Evolution of the average north polar field (solid line) and average south polar field (dotted line) in G for latitudes above $80^\circ$. 
surface. As with the open flux the total dipole is less than before. In Figure 4(c) the variation of the average polar flux is shown. Since the initial distribution assumes there to be a balance between meridional flow and supergranular diffusion, when meridional flow is switched off the polar fields immediately start to diffuse towards the equator. However, within a cycle an equilibrium variation of the polar fields is set up and they oscillate between the values of $\pm 3$ G in each hemisphere. The cyclic switching of the polar fields is still maintained, but, the polar fields are now much weaker, since it is harder for new emerging flux to reach the poles. Simulations have also been run where the initial flux distribution is consistent with there being no meridional flow and exactly the same results are found.

4.3. LONGITUDINAL AVERAGES OF SURFACE MAGNETIC FIELD

To consider why there is a different phase relationship between the open and surface flux when meridional flow is absent, the differences in the surface distributions of the radial magnetic field are considered. This can be seen in Figure 5, where the longitudinally averaged radial magnetic field at the solar surface is shown for the first 4 solar cycles as a function of time in the simulation and sine-latitude. The time resolution of the plots is 27 days. By expressing the average in terms of sine-latitude this emphasis's the lower latitude fields compared to the polar fields in each hemisphere. For each plot the field is set to saturate at 5 G. Figure 5(a) shows the results for all three surface effects, while Figure 5(b) is for meridional flow switched off. As can be seen there are strong differences in the surface pattern as a result of meridional flow.

In both simulations the leading polarity flux in each hemisphere, which is of the same sign as the polar field, lies at lower latitudes than the trailing polarity flux (which is of opposite sign to the polar fields). Due to this latitudinal distribution, the trailing polarity flux can preferentially reach the poles and cancel the polar field. It then builds up a new polar field (of trailing polarity). In the next cycle the process repeats itself.

When meridional flow is included (Figure 5(a)) the trailing polarity flux can be seen to head poleward in a series of discrete ‘poleward surges’ (Wang, Nash, and Sheeley, 1989b). These surges are mainly due to the effect of meridional flow on new flux emergences. The surges enhance the latitudinal separation of the leading and trailing polarity flux and aid first, the cancellation and then the formation of a concentrated polar cap. The surges or ‘herringbone pattern’ are also seen in Kitt Peak and Wilcox Solar Observatory observations (Figure 7(b) of van Ballegooijen, Cartledge, and Priest, 1998; Figure 6 of Wang, Lean, and Sheeley, 2000; Figure 1 of Gaizauskas, Mackay, and Harvey, 2001). Comparison of Figure 5(a) with these observations shows that the slope of the poleward surges with respect to time is roughly correct, so that the relative values of diffusion and meridional flow in the simulation are likely to be realistic.
At solar minimum the radial magnetic field at the solar surface has a top-knot profile produced by meridional flow pushing flux poleward (Wang and Sheeley, 1991; Wang, Nash, and Sheeley, 1989b). At this phase in the cycle the polar regions are the most concentrated and unipolar. With this profile the lower order harmonics at the solar surface dominate. These harmonics, due to their height dependence (Equation (5)), contribute significantly to the open magnetic flux even though the total surface flux is much less than at maximum (Figure 3). In contrast, at solar maximum the flux is now located in strong concentrations at low latitudes where opposite polarities are closely intermingled. With this the higher order harmonics now become dominant and the lower order ones weak. Therefore, even though there is much more surface flux at solar maximum, the open flux is small since the higher order harmonics fall off fast with height, leaving only the weak lower order harmonics to contribute to the open flux at the source surface. Therefore, even though there is much more flux at solar maximum due to the potential assumption which fixes how individual harmonics fall off with height, the net open flux is less
than at minimum. Meridional flow therefore plays a dominant role on the phase relationship of the open flux by pushing the trailing polarity flux poleward in each hemisphere on a time scale of \( \tau_m = R_\odot / u_0 = 2 \) years and altering on this time scale the surface distribution from higher to lower order harmonics.

A very different surface pattern is seen when meridional flow is absent. Now there are no poleward surges and the only way the trailing polarity flux can extend poleward is through diffusion (\( \tau_d = \frac{R_\odot^2}{D} = 26 \) years). This is ineffective in pushing flux poleward and only weak polar fields (and lowest order harmonics) are produced at cycle minimum. The surface field now has a mid-latitude bipolar structure which can clearly be seen to be the strongest after cycle maximum. When meridional flow is present, it stops this mid-latitude bipolar structure from forming and instead produces a polar top knot profile.

When meridional flow is not included, the maximum in open flux occurs in the manner previously described by Wang, Sheeley, and Lean (2000). At maximum, the open flux is limited due to the fact that the surface field is very intermingled. The higher order harmonics which fall off rapidly with height (Equation (5)) dominate over lower ones. As the flux cancels, the leader and follower flux (which lie at different latitudes) form a concentrated mid-latitude bipolar structure. The mid-latitude bipolar structure can form, since in the absence of meridional flow opposite magnetic polarities are not separated in latitude (Wang, Nash, and Sheeley, 1989). In doing so, the higher order harmonics start to decay and shift slightly towards the lower order ones. In the first few years after cycle maximum, the surface fields with bipolar form are strong since, when meridional flow is absent, the flux is not being pushed to higher latitudes where the gradient in differential rotation increases and diffusion becomes more effective (Mackay, Priest, and Lockwood, 2002). This leads to an increase in the amount of open magnetic flux. The magnetic flux continues to be sheared by differential rotation at the latitude it emerges and cancels \textit{in situ}. Since flux is not pushed poleward, the lowest order harmonics (which are present in a top-knot profile and contribute significantly to the open flux even when the surface flux is weak) never form. Therefore, as the surface flux decays at mid-latitudes the open flux decays because the flux pattern is not shifting to the lowest order harmonics fast enough to maintain the open flux. The maximum in the open flux occurs when there is an optimal balance between the strength of the surface fields and the order of the harmonics used to describe it. In the absence of meridional flow and with the value of diffusion coefficient used, this occurs a few years after cycle maximum.

From Figure 5 it is clear that even though the absence of meridional flow does produce the correct phase relationship between the open and surface flux, it does so for the wrong reasons by not allowing fields to be pushed poleward and producing polar caps dominated by the lowest order harmonics. Two extreme cases have been considered with respect to the relative values of meridional flow and supergranular diffusion. In the next section the full parameter space of these quantities and others are considered to determine if realistic variations of present values can reproduce
Figure 6. Graph showing the regions of meridional flow/diffusion parameter space where the correct phase relationship between surface and open flux is obtained (area shaded with solid lines) and where the correct surface flux distribution is obtained (area shaded with dotted lines). The ‘X’ marks the combination used in Section 4 and in other studies.

the correct surface pattern and phase relationship between the open and surface flux.

4.4. PARAMETER SPACE

In flux transport simulations there are a large number of parameters which may be varied. Some directly relate to the input data of new bipolar magnetic regions, while others relate to the flux transport model. These parameters are now varied to determine what effect, if any, they have on the phase relationship of the surface and open flux.

To begin with, the initial profile of the magnetic field is changed to be that of Figure 2(b), but this produces no appreciable change from the results in Figures 3 and 4. The input data is now considered. From the observations of Wang and Sheeley (1989) we are happy that the butterfly diagram, relative number of positive, negative and zero tilt angle bipoles along with their latitudinal variation are correct. These quantities are left unchanged. Figure 1(b) shows the emergence rate of bipoles which can be related to the sunspot number. The sunspot number has been considered for hundreds of years and may show significant variations from one cycle to the next (see Figure 3 in Lockwood, Stamper, and Wild, 1999). In the previous simulations the emergence rate was constant throughout the simulations and so roughly equal amounts of flux emerged in each cycle. To determine whether variations in the amount of flux emerging from one cycle to the next can affect the phase relationship of surface to open flux, simulations are run were the emergence rate of one cycle is increased/decreased relative to the others (with all three surface effects included). Varying the amount of flux in each cycle produced no
difference in the phase relationship, as the maximum in open flux still occurred at solar minimum when the polar caps produced by meridional flow are the strongest. Simulations have also been carried out where the length and overlap period varies from cycle to cycle, but it was found that this also has no effect on the phase relation of the surface to open flux.

It is clear that meridional flow plays a significant role in the variation of the open flux. The magnitude of the diffusion coefficient and meridional flow are now varied in the flux transport model to determine if a combination (which fits observed values) can be found to reproduce the observed phase relationship of surface and open flux. In Figure 6 the graph shows the results, where the $x$-axis represents meridional flow ($\text{m s}^{-1}$) and the $y$-axis the diffusion coefficient ($\text{km}^2 \text{s}^{-1}$). Due to computational requirements in scanning the parameter space a low resolution of the diffusion coefficient (200 $\text{km}^2 \text{s}^{-1}$) and meridional flow (3 $\text{ms}^{-1}$) is used. Even with this low resolution the result is clear. The vertical lines show the region of parameter space where the open magnetic flux lags behind the surface flux by the correct period. This occurs when diffusion dominates over meridional flow. If, however, the diffusion coefficient becomes too large, the open flux falls into phase with the surface flux. In contrast, the dotted lines show the combinations where realistic surface flux distributions are produced. The ‘X’ denotes the values used in Section 4.1. From this it can be seen that no realistic combination of meridional flow and diffusion coefficient in the flux transport model can be found (under potential coronal fields) that produces the correct relationship between the open and surface flux.

5. Discussion

In this paper the origin and evolution of the Sun’s open magnetic flux has been considered over many solar cycles through magnetic flux transport simulations. The simulations describe the evolution of the radial magnetic field at the surface of the Sun as new magnetic flux emerges and is pushed poleward (Wang, Nash, and Sheeley, 1989b). At regular intervals, potential magnetic fields with a source surface are constructed to determine the amount of open magnetic flux. The simulations follow the surface and open magnetic flux for seven 11-year solar cycles, where in each cycle roughly 2100 bipolar magnetic regions emerge at random longitudes. Due to the work of Wang, Sheeley, and Lean (2000) and Mackay, Priest, and Lockwood (2002), the latitude of emergence, relative number of positive, negative and zero tilt angles, along with the latitudinal variation of tilt angle are fitted as accurately as possible to observations (Wang and Sheeley, 1989). The flux transport model and input data produced an accurate evolution of the surface distribution of the magnetic field of the Sun.

In the simulations, as with previous studies (Wang, Sheeley, and Lean, 2000; Mackay, Priest, and Lockwood, 2002), the open magnetic flux evolves indepen-
OPEN MAGNETIC FLUX

In all simulations the north–south axisymmetric dipole component at the source surface dominates over the east–west non-axisymmetric dipole component for the majority of the simulation except for a short period around solar maximum. In previous papers (Wang, Sheeley, and Lean, 2000; Mackay, Priest, and Lockwood, 2002) differential rotation and diffusion were the key physical effects which produced the dominant variations in the open flux as either one or two bipolar magnetic regions evolved. However, in full solar cycle simulations, which are much more complex and have polar magnetic fields and multipole bipoles, the effect of differential rotation found for a single bipole is averaged out by neighboring bipole interactions. Instead, meridional flow now plays the dominant role in the variation of open flux by pushing magnetic flux towards the poles.

When all three surface effects are included with values derived from observations, the surface and open magnetic flux are $\pi$ out of phase. The maximum in open flux occurs at solar minimum, which is inconsistent with observations (Lockwood, Stamper, and Wild 1999; Wang, Lean, and Sheeley, 2000). Only when meridional flow is negligible compared to diffusion is the correct phase relationship obtained. This occurs because, when meridional flow is absent, a mid-latitude bipolar structure may be formed 1–2 years after cycle maximum (Wang, Sheeley, and Lean, 2000), which then gives the desired effect of peaking the open flux at the observed time. When meridional flow is present (with a time scale of 2 years) the mid-latitude bipolar structure is destroyed and a polar one formed instead. Although in the absence of meridional flow the correct phase relationship may be obtained, it occurs for unrealistic surface distributions. It is concluded that there is no reasonable combination of flux transport parameters that can produce the observational variation in open magnetic flux when the present potential model is used to determine the magnitude of the open magnetic flux.

Although the present simulations do not explain the origin of the observed lag between the open and surface magnetic flux they do raise a number of questions on the methods used to model it. The procedure uses two distinct models. The first is the flux transport model, which evolves the radial magnetic field at the surface of the Sun. The second is the potential source surface model, which takes the computed radial magnetic field and constructs a potential coronal magnetic field from it. The limitations and improvements required from both models are now considered.

To begin with, consider the magnetic flux transport model. Flux transport models have been used for many years (DeVore, Sheeley, and Boris, 1984; Sheeley, DeVore, and Boris, 1985; Sheeley, Nash, and Wang, 1987; Wang, Nash, and Sheeley, 1989; van Ballegooijen, Cartledge, and Priest, 1998; van Ballegooijen, Priest, and Mackay, 2000; Mackay, Gaizauskas, and van Ballegooijen, 2000; Mackay, Priest, and Lockwood, 2002) with great success in modeling the evolution of the radial magnetic field of the Sun. The values used to simulate the three main physical effects (differential rotation, meridional flow and supergranular diffusion) have
been fitted to observed values and well tested. One area of improvement would be to have a meridional flow velocity which is time dependent such as observed by Hathaway (1996). In the present simulations we are using a lower bound and any change would involve increasing the magnitude of meridional flow at certain times in the simulation (therefore decreasing the time scale over which it acts). This would break up the mid-latitude bipolar structure faster and have the opposite effect to what is required, as demonstrated in Figure 6. The diffusion coefficient in the present simulations is uniform and derived from observations taken at mid-latitudes. If it is in fact spatially varying with the value changing towards the poles, this may alter the results. At the present time, there is, however, no observational evidence to support this. Detailed observations of the polar regions would be required to clarify the issue.

In recent years, MDI has shown that there are continual small scale changes to the magnetic field of the Sun, the so-called ‘Magnetic Carpet’. Such small scale emergences and cancellations are not considered here. To include them the number of harmonics used to describe the field would have to be drastically increased from its present value. Increasing the number of harmonics along with the potential model would not change the present variation of the open flux due to the fast fall off of the harmonics with height. We believe with present observational evidence that the magnetic flux transport model provides an accurate evolution of the large scale magnetic field of the Sun. Under the potential assumption it is this large scale field which is already being modeled that produces the dominant contributions to the open flux.

To determine the amount of open magnetic flux a potential source surface model is used. This model assumes that in the corona there are no electric currents or free magnetic energy and that the field becomes radial at $2.5 \, R_\odot$. It stipulates how the harmonics used to describe the surface magnetic field vary with height (Equation (5)). The higher the harmonic the faster its amplitude decreases with height. In the simulations meridional flow pushes flux poleward to produce unipolar polar caps and reduces the magnetic field to lower harmonics. Hence with the harmonics height dependence it is not surprising that the maximum in open flux occurs at solar minimum. Since the potential model produces this behaviour for the harmonics, it suggests that it may not contain the correct physics to describe the evolution of the open flux over a complete solar cycle. Therefore, to determine the origin and evolution of the open flux, it seems that improvements in the coronal model are required.

Such improvements may include evolving the surface and corona magnetic fields together (van Ballegooijen, Priest, and Mackay, 2000; Mackay, Gaizauskas, and van Ballegooijen, 2000) to follow the response of the coronal field as the surface field is evolved. Such a method following the expansion of flux tubes as they are stressed may produce a completely different dependence of the surface harmonics with height, especially around the time of solar maximum. In such a model, the role of the magnetic carpet may become important and a much more
realistic evolution of both the closed and open magnetic field lines could be ob-
tained. Presently such a simulation which would also have to take into account the
loss of energy and helicity due to eruptions (over many solar cycles) is beyond
computing resources. Secondly, the solar wind may have to be included. This will
produce a much more realistic description of how magnetic field lines are blown
open compared to the presently used source surface model.

The conclusion of the simulations is that more detailed physics is required
within the coronal model if the correct phase relationship between the open and
surface flux is to be obtained. This, however, is inconsistent with the results of
Wang, Lean, and Sheeley (2000), who used a similar potential model along with
synoptic magnetograms from Mount Wilson and Wilcox Solar Observatories to
derive the same phase relationship as Lockwood, Stamper, and Wild (1999). The
disagreement between Wang, Sheeley, and Lean (2000) and this paper may be a
result of a difference in strength of the polar magnetic fields observed at solar
minimum and those produced by the model. In the model the polar fields are
formed in a manner consistent with the amount of flux emerging (Figure 1), rate of
poleward transport (Equation (2)) and the magnitude of the diffusion coefficient.
These parameters are derived from observed values and produce the polar fields in
a self-consistent way. In contrast, in magnetogram data the polar fields cannot be
observed with any certainty and therefore when the potential model is constructed
the lower order harmonics (which dominate around minimum) are not observed.
A full or partial correction for this may be inherent in the ‘line saturation’ fac-
tor employed by Wang, Lean, and Sheeley (2000), which is a strong function of
latitude (Ulrich, 1992; Wang and Sheeley, 1995). However, since the polar fields
cannot be observed with any certainty, it is not clear that a correction factor can
reproduce the true polar field. Therefore, the magnetograph data may even after
the correction have an enhanced mid-latitude field compared to the polar fields as
occurs in Section 4.2. To resolve this issue and determine whether the polar fields
formed in the model are correct detailed observations of the polar field would be
required which are not available at the present time. This is beyond the scope of
the present paper.

In this paper, we have considered how the Sun’s open magnetic flux varies with
flux transport and potential source surface models. We were unable to produce the
observed lag between the surface and open flux with a realistic balance of diffusion
and meridional flow such that the surface field evolution is accurately modeled.
From this it is suggested that the potential extrapolations of the surface field prob-
ably do not contain the correct physics to describe the magnitude variation of the
open flux. Further observational investigation and more detailed modeling of the
evolution of the Sun’s open magnetic flux, especially with respect to the polar and
coronal fields, will be required before we are able to understand its origin and
variation.
Acknowledgements

The authors would like to thank Aad van Ballegooijen, who initially developed the codes used in this paper, and the UK Particle Physics and Astronomy Research Council for financial support.

References


