

# The solar magnetic field since 1700

## I. Characteristics of sunspot group emergence and reconstruction of the butterfly diagram

J. Jiang, R. H. Cameron, D. Schmitt, and M. Schüssler

Max-Planck-Institut für Sonnensystemforschung, 37191 Katlenburg-Lindau, Germany

Received 18 November 2010 / Accepted 31 January 2011

### ABSTRACT

We use the historic record of sunspot groups compiled by the Royal Greenwich Observatory together with the sunspot number to derive the dependence of the statistical properties of sunspot emergence on the cycle phase and strength. In particular we discuss the latitude, longitude, area and tilt angle of sunspot groups as functions of the cycle strength and of time during the solar cycle.

Using these empirical characteristics the time-latitude diagram of sunspot group emergence (butterfly diagram) is reconstructed from 1700 onward on the basis of the Wolf and group sunspot numbers. This reconstruction will be useful in studies of the long-term evolution of the Sun's magnetic field.

**Key words.** sunspots – Sun: surface magnetism – Sun: dynamo

### 1. Introduction

The synoptic record of sunspot emergence is an important input into, for example, long term reconstructions of the solar open flux (e.g. Cameron et al. 2010), solar irradiance variations (e.g. Crouch et al. 2008) and is relevant for understanding the solar dynamo (for a recent review see Charbonneau 2010). The quantities which are used in these types of studies include the sunspot areas, emergence latitudes and longitudes as well as the tilt angle between sunspots of opposite polarities within a group.

As an example of such a study, in Cameron et al. (2010) we used the observed Royal Greenwich Observatory<sup>1</sup> (RGO) records of sunspot areas, longitudes, latitudes and the Mount Wilson Observatory and Kodaikanal records of tilt angles (Howard et al. 1984, 1999; Sivaraman et al. 1993) as the input to a surface flux transport model (e.g. Devore et al. 1985; Sheeley et al. 1985; Wang et al. 1989; Baumann et al. 2004). The results from the model was the large-scale evolution of the surface magnetic field, which was extrapolated into interplanetary space using a current sheet source surface (CSSS) extrapolation (e.g. Zhao & Hoeksema 1995). The open field calculated from the model was then compared to that inferred from observations of the geomagnetic aa index. The time period analyzed was restricted by the fact that the RGO dataset only extends back to 1874 and the MWO and Kodaikanal tilt angle datasets are even shorter.

The purpose of this paper is to construct partially synthetic datasets of sunspot emergence covering the period from 1700 to 2010. In the second paper in this series we intend to use these semi-synthetic records with the surface flux transport model and CSSS extrapolation. The semi-synthetic data sets however have a much wider application, for example in irradiance studies and in understanding the solar dynamo. We therefore here present the analysis and methods for creating them.

The time dependence of the emergence is taken from the sunspot number data, either the monthly group sunspot number,  $R_G$  (Hoyt & Schatten 1998), or the monthly Wolf sunspot number,  $R_Z$  (Wolf 1861). Correlations between the strength of the cycle, derived from  $R_G$  or  $R_Z$ , and the areas, emergence latitudes and longitudes, and tilt angles of sunspot groups are sought. For this purpose we use the RGO record of sunspot group areas, latitudes, longitudes as well as the MWO and Kodaikanal records of sunspot group tilt angles. Unlike previous studies (e.g. Li et al. 2003; Solanki et al. 2008; Dasi-Espuig et al. 2010; Ivanov & Miletsky 2011), we consider correlations of many of the emergence properties with cycle properties derived from the monthly sunspot number.

These correlations are then used in conjunction with the  $R_G$  and  $R_Z$  records to construct artificial sunspot group data extending back to 1700. As the time dependence is taken from observations and the other properties of the sunspot groups are synthetic, the constructed timeseries is semi-synthetic. Since the correlations are only statistical in nature, the individual reconstructions are realizations drawn randomly from a population with the observed statistics. This enables Monte-Carlo-type studies on the longer term evolution of, e.g. the Sun's open flux, polar fields and irradiance variations.

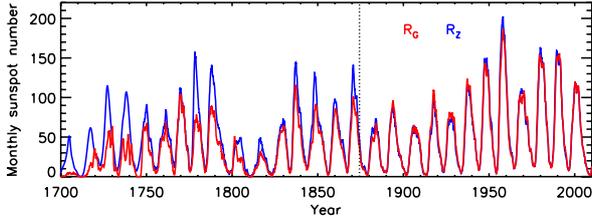
The paper is structured as follows: in Sect. 2 we describe the datasets and ways to define various cycle parameters such as cycle strength, starting time and length of each cycle. In Sect. 3 we discuss the correlations between the spatial distribution and the properties of the cycle as determined from  $R_G$ . In Sect. 4 we use these correlations to reconstruct the butterfly diagram from 1700 onwards.

### 2. Cycle parameters determined from sunspot numbers

#### 2.1. Sunspot number datasets

Monthly values of the Wolf sunspot number,  $R_Z$ , are available from 1749 onward and yearly values reach back to 1700.

<sup>1</sup> All data was obtained from the NOAA website <http://www.ngdc.noaa.gov/stp/solar/solardataservices.html>



**Fig. 1.** Time evolution of monthly Wolf sunspot number  $R_Z$  (blue line) and group sunspot number  $R_G$  (red line). The dotted vertical line denotes the year 1874, after which the two sunspot number datasets are nearly identical. Note that  $R_G$  ends in 1995, after which it is assumed to be equal to  $R_Z$ .

Monthly values of  $R_Z$  before 1749 can be estimated by interpolation of yearly  $R_Z$  values. The group sunspot number  $R_G$  extends further back in time, to 1610, and is again interpolated to obtain monthly values when the dataset is incomplete. As described by Usoskin (2008) and Hathaway (2010) the two datasets have different definitions and depend on different combinations of solar observations.

Figure 1 shows the two sunspot numbers over the period from 1700 to 2010. After the 1870s they are nearly identical. Since this covers the period of more detailed RGO sunspot data, the two data sets are almost equivalent for use in determining the empirical correlations with the RGO data. We have (arbitrarily) chosen to use  $R_G$ . Differences however are to be expected in the reconstructed sunspot group data prior to 1874, as will be seen in Sect. 4.

## 2.2. Cycle parameters

On the basis of the  $R_G$  data we define three parameters for each cycle. Two of these parameters concern the strength of a cycle. The first,  $S_n$ , is the maximum of the 12 month running mean of  $R_G$ . The second,  $\bar{S}_n$ , is the sum of  $R_G$  over the cycle.

We also need information as to the timing of the cycle, and for this we use the time of the solar minima,  $t_{\min}$  (Harvey & White 1999), which we take from the NGDC website. These times and cycle strengths are listed in Table 1.

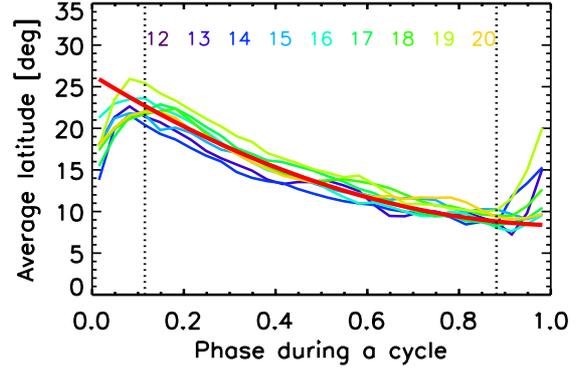
## 3. Characteristics of sunspot group emergence derived from the RGO sunspot data

In this section we discuss the empirical relationships between the strength of the cycle given in Table 1 and the latitudes, longitudes and areas of sunspot groups as recorded in the RGO dataset. The RGO records cover the period from 1874 to 1976 (cycles 12 to 20). The sunspot groups are considered at the times of their maximum reported area.

Correlations between the strength of the cycle and the tilt angle of sunspot groups rely on the MWO and Kodaikanal data as discussed in Dasi-Espuig et al. (2010). Some properties, such as the average latitude at which sunspots emerge, vary throughout the solar cycle. In this case it is important to consider the data from different cycles at the same phase. We then look for correlations at a fixed phase through the cycle, where the cycle is taken to begin and end at adjacent activity minima.

### 3.1. Latitude distribution

The latitude distribution of sunspots is the clearest example of where it is necessary to consider the phase during the cycle: early



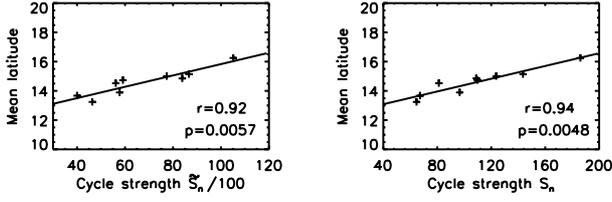
**Fig. 2.** Average latitude of sunspot groups at different cycle phases. Colors indicate different cycles (cycles 12–20). The thick red curve shows the polynomial fit given by Eq. (1) to all cycles. The two vertical dotted lines indicate the times when the overlapping of cycles strongly affects the average latitudes.

**Table 1.** Parameters for solar cycles  $-4$  to  $23$  (from 1700 onward) derived from the group sunspot number  $R_G$ .

Cycle No.	$t_{\min}$	$\bar{S}_n/100$	$S_n$
-4	1698.0	3	5
-3	1712.0	16	35
-2	1723.5	33	64
-1	1734.0	31	53
0	1745.0	36	67
1	1755.2	43	70
2	1766.5	61	103
3	1775.5	45	78
4	1784.7	70	89
5	1798.3	23	51
6	1810.6	19	32
7	1823.3	41	64
8	1833.9	60	116
9	1843.5	64	91
10	1856.0	55	85
11	1867.2	57	101
12	1878.9	40	67
13	1889.6	57	96
14	1901.7	46	64
15	1913.6	59	109
16	1923.6	56	81
17	1933.8	77	123
18	1944.2	86	143
19	1954.3	105	186
20	1964.9	83	108
21	1976.5	100	154
22	1986.8	92	156
23	1996.4	81	119

in the cycle sunspots appear at higher latitudes than later in the cycle.

We break the time between adjacent minima into 30 equal phases. For each cycle,  $n$ , we can then calculate the mean latitude,  $\lambda_n^i$ , averaged over the  $i$ th phase bin. Figure 2 shows  $\lambda_n^i$  for cycles 12 to 20 as a function of the phase of the cycle. Because the cycles partially overlap, the first 3 phase bins near the start and the last 3 near the end of a cycle show a mixture of spots from adjacent cycles. This mixture is of lower-latitude spots from the end of a cycle and higher-latitude spots from the beginning of the subsequent cycle. The average latitude during these initial and final phases is thus difficult to interpret.



**Fig. 3.** Correlation between cycle averaged latitudes  $\lambda_n$  and cycle strength defined by the total sunspot number ( $\tilde{S}_n$ , left panel) and the maximum sunspot number ( $S_n$ , right panel), respectively.

The phase average  $\lambda^i$  over all cycles defined as  $\frac{1}{9} \sum_{n=12}^{20} \lambda_n^i$  is well fit by a second degree polynomial:

$$\lambda^i = 26.4 - 34.2(i/30) + 16.1(i/30)^2 \quad (1)$$

over the range  $4 \leq i \leq 27$ . The rms difference between the fit and the mean latitudes is  $0.3^\circ$ .

As previously reported (Li et al. 2003; Solanki et al. 2008), there is a strong correlation between the strength of the cycle and the average latitudes of emergence. To evaluate this correlation we calculated the mean latitude of emergence as

$$\lambda_n = \frac{1}{24} \sum_{i=4}^{27} \lambda_n^i. \quad (2)$$

Figure 3 shows the relation between the average latitudes  $\lambda_n$  and cycle strengths defined by  $\tilde{S}_n$  (left panel) and  $S_n$  (right panel). In both cases, a statistically significant ( $p$ -value  $< 0.05$ ) correlation coefficient higher than 0.9 is found. The correlations between  $S_n$  and  $\lambda_n$  ( $r = 0.94$ ) and between  $\tilde{S}_n$  and  $\lambda_n$  ( $r = 0.92$ ) are similar. We hereafter focus on  $S_n$ . The linear fit for the  $\lambda_n$  is given by

$$\lambda_n = 12.2 + 0.022S_n. \quad (3)$$

This observed correlation can be used with Eq. (2) to model the phase dependence of the mean latitude of emergence for different cycles:

$$\lambda_n^i = (26.4 - 34.2(i/30) + 16.1(i/30)^2)(\lambda_n / \langle \lambda_n \rangle_{12-20}) \quad (4)$$

for  $1 \leq i \leq 30$  and where  $\langle \lambda \rangle_{12-20} = 14.6^\circ$  is the average latitude of sunspot emergence over all the cycles. The fit to each cycle is shown in Fig. 4. The mean rms deviation between observation and reconstruction, excluding the first and last two years of each cycle, is  $1.33^\circ$ .

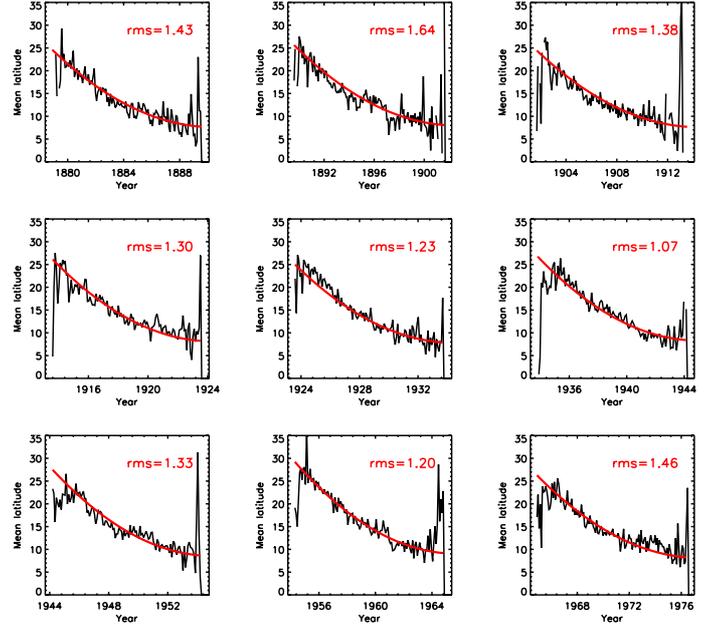
### 3.2. Width of the latitude distribution

Sunspots emerge over a range of latitudes at any phase of the cycle. We consider the standard deviation,  $\sigma_n^i$ , of the latitudinal distribution during phase  $i$  of cycle  $n$ . The upper panel in Fig. 5 shows  $\sigma_n^i$  for cycles  $n = 12-20$ . A tighter relationship is found if we consider the ratio  $\sigma_n^i / \lambda_n^i$  which is shown in the lower panel of the same figure.

A second-order polynomial fit

$$\sigma^i = (0.14 + 1.05(i/30) - 0.78(i/30)^2)\lambda^i \quad (5)$$

matches the data well. In our semi-synthetic reconstructions we assume a Gaussian distribution with a half width of  $\sigma_n^i$  and exclude points deviating from the mean by more than  $2\sigma_n^i$ .



**Fig. 4.** Comparison of the mean latitudes of sunspot emergence for cycles 12–20 between the observation (black line, monthly average) and the fit (red line). The rms deviations (in degrees) between observation and reconstruction are given excluding the first and the last two years of a cycle.

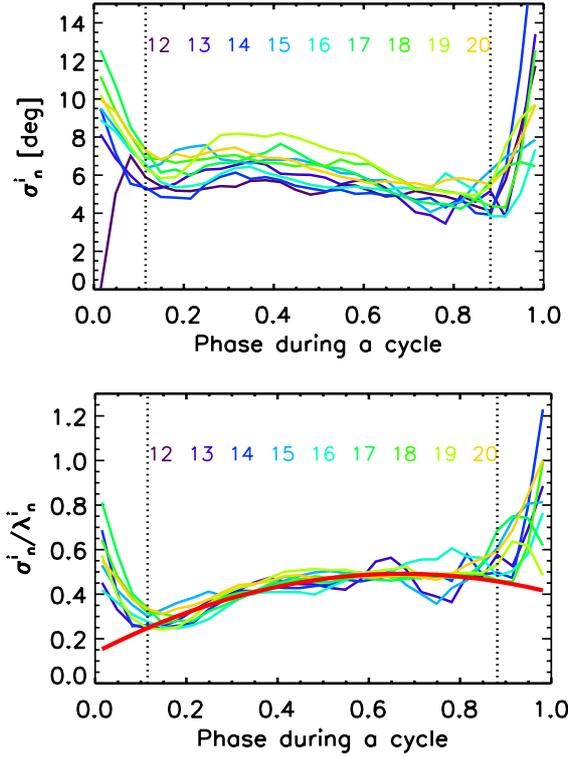
### 3.3. Longitude distribution

The emergence longitudes of sunspot groups is known to be not entirely random (e.g. Bumba & Howard 1965; Bai 1988; Berdyugina & Usoskin 2003; Zhang et al. 2007). Castenmiller et al. (1986) introduced the term “sunspot nests” to describe the tendency for sunspots to appear near where other spots. Up to 30% of the sunspot groups have previously been found to be associated with such nests.

Our motivation for considering the longitudes distribution is, for example, that the open flux of the Sun during activity maxima is dominated by the equatorial low order multipoles (see, e.g. Cameron et al. 2010). The strengths of these multipoles, and the dipole in particular, depend on how randomly or systematically the sunspots appear in longitude. We therefore begin by considering the equatorial dipole moment of an individual sunspot group. The first step is to convert the sunspot area,  $A$ , in the RGO dataset into magnetic flux. Here we follow van Ballegooijen et al. (1998) and Baumann et al. (2004) and take the total flux of the group to be proportional to the area of the active region (sunspot area and plage area)

$$A_R = A + 414 + 21A - 0.0036A^2 \quad (6)$$

where all values are in  $\mu\text{Hem}$  (Chapman et al. 1997). The equatorial dipole moment of the sunspot group is then assumed to be proportional to the area of the group,  $A_R$ , multiplied by the separation between the opposing polarities which we take to be proportional to  $A_R^{1/2}$ . For a sunspot group  $j$  in the northern hemisphere and near the equator, with area  $A_{R,j}$  and central longitude  $\phi_j$ , the axis of the equatorial dipole is orientated in the direction  $\phi_j \pm 90^\circ$ . The component of the dipole in the equatorial plane in direction  $\phi$  is thus proportional to  $A_{R,j}^{3/2} \cos(\phi - \phi_j)$ . For a number of sunspots groups in the northern hemisphere, the



**Fig. 5.** Upper panel: standard deviation  $\sigma_n^i$  of the latitudinal distribution of phase-binned latitudes for cycles 12–20. The line style is the same as in Fig. 2. Lower panel: similar except  $\sigma_n^i/\lambda_n^i$  is shown. The thick red curve shows the polynomial fit given by Eq. (5).

component of the resulting dipole moment in the direction  $\phi$  is proportional to

$$\sum_{\text{north}} A_{R,j}^{3/2} \cos(\phi - \phi_j). \quad (7)$$

Since the sunspot groups in the southern hemisphere have the opposite polarity orientation, the corresponding component of the dipole moment is proportional to

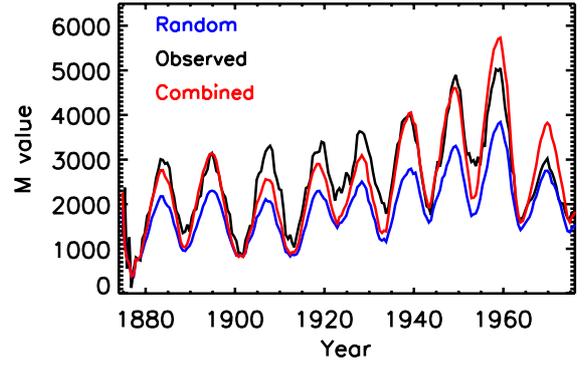
$$- \sum_{\text{south}} A_{R,j}^{3/2} \cos(\phi - \phi_j). \quad (8)$$

Therefore dipole moment of all the spot groups from both hemispheres is proportional to

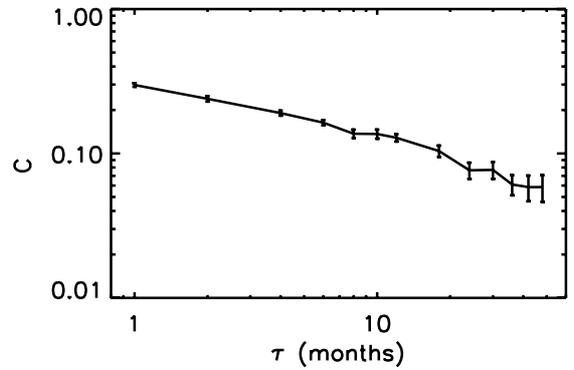
$$m = \sum_{\text{north}} A_{R,j}^{3/2} \cos(\phi - \phi_j) - \sum_{\text{south}} A_{R,j}^{3/2} \cos(\phi - \phi_j). \quad (9)$$

Clearly,  $m$  depends on the sunspots which are included in the sum as well as the direction  $\phi$ . We define  $m(t, \tau, \phi)$  to include in the sum all spots emerging between times  $t$  and  $t + \tau$ .

We are not able to reproduce the longitudes at which sunspots have appeared from 1700 onwards or even those longitudes where nesting has occurred. Our aim is to obtain statistical information about the degree of nesting by measuring the degree of non-randomness present in the RGO data. We do this by creating three copies of the RGO records. In the first copy we replace the observed longitudes with randomly chosen longitudes from a uniform distribution. We use the subscript notation  $m_{\text{ran}}$  for this dataset. The second copy has the longitudes of its spots changed so that they all appear at  $0^\circ$  in the northern hemisphere and  $180^\circ$  in the southern hemisphere. This choice maximizes the



**Fig. 6.** Time evolution of the proxy for the equatorial dipole moment,  $M$ , during 1874–1976 for  $\tau = 6$  months. The random (blue curve), observed (black curve) and combined (red curve) models for the longitude distribution of sunspot emergence are shown. The value of  $c = 0.15$  corresponds to  $\tau = 6$  months.



**Fig. 7.** Degree of non-randomness,  $c$ , as a function of the period over which we take the sunspots,  $\tau$ . Error bars represent the standard deviation resulting from 20 sets of random longitudes to obtain  $M_{\text{ran}}$ .

equatorial magnetic dipole moment since sunspot groups in opposite hemispheres have opposite polarity orientations in accordance with Hale’s law. We use the subscript notation  $m_{\text{ord}}$  for this dataset. The third copy retains the observed longitudes and uses the subscript notation  $m_{\text{obs}}$ .

For a given  $\tau$  we measure the amount of nonrandomness,  $c(\tau)$ . We assume that the magnitude of the observed sunspot dipole moment

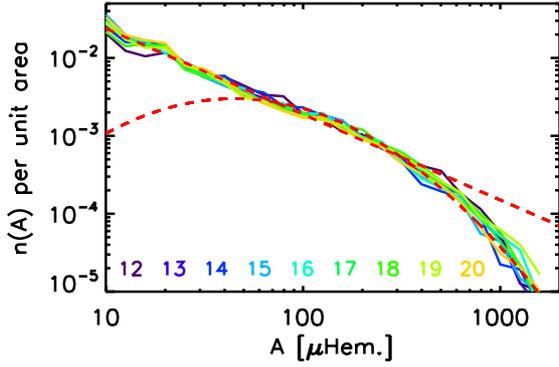
$$M_{\text{obs}}(t, \tau) \equiv \max_{\phi} \{m_{\text{obs}}(t, \tau, \phi)\} \quad (10)$$

is the magnitude of a linear combination of the random and ordered datasets

$$M_c(t, \tau) \equiv \max_{\phi} \{c(\tau)m_{\text{ord}}(t, \tau) + [1 - c(\tau)]m_{\text{ran}}(t, \tau)\} \quad (11)$$

and seek the  $c(\tau)$  which minimizes the rms differences. Figure 6 shows  $M_{\text{obs}}$ ,  $M_{\text{ran}}$  and  $M_c$  for the example  $\tau = 6$  months. The difference between  $M_{\text{obs}}$  and  $M_{\text{ran}}$  during activity maxima reflects the amount of nesting, i.e., the nonrandom longitude distribution. The similarity of all the curves near the minima is a consequence of the fact that, when there are few spot groups, they are automatically highly ordered. We show the dependence of  $c$  on  $\tau$  in Fig. 7.

For different studies the appropriate value of  $\tau$  will vary. For irradiance studies  $\tau = 1$  month would seem to be an appropriate choice because it is the instantaneous clustering of the sunspots which is important. For surface flux transport simulations,  $\tau = 6$  months is more relevant as this is approximately the



**Fig. 8.** Number density function of the group areas for each cycle of RGO area dataset. The two red dashed curves show the fits from Eqs. (12) and (13).

time it takes for the emerging flux to be sheared by differential rotation (Schrijver & Zwaan 2000, p. 162).

### 3.4. Area distribution

In this section we consider the area distribution of sunspot groups based upon the RGO dataset. As previously stated, we consider each group only at the time when it has its maximum recorded area. We begin by considering the area distribution of the entire dataset. We will then look at the partially related questions of its dependence on the phase during the cycle and on the latitudes at which the spots appear.

#### 3.4.1. Area distribution function

Figure 8 shows the number density function of sunspot group areas. The behaviour is approximately a power law below  $300 \mu\text{Hem}$  with a turnover to an almost log-normal distribution above (see also Zhang et al. 2010). There is a large range (from  $60$  to  $300 \mu\text{Hem}$ ) where both functional forms are good approximations to the data. The fits for the two sections of the curves are

$$n(A) = 0.3A^{-1.1} \text{ for } A < 300 \mu\text{Hem} \quad (12)$$

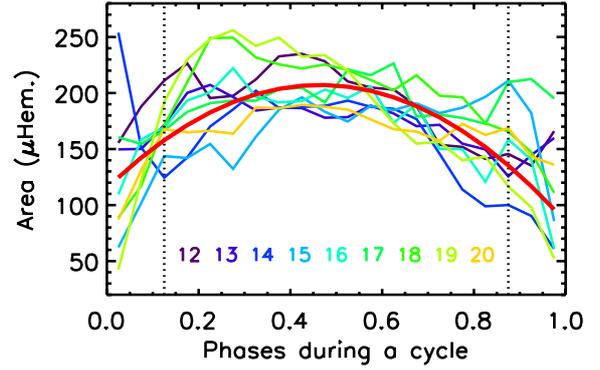
and

$$n(A) = 0.003 \exp\left[-\frac{1}{2 \ln 3} (\ln A - \ln 45)^2\right] \text{ for } A > 60 \mu\text{Hem}. \quad (13)$$

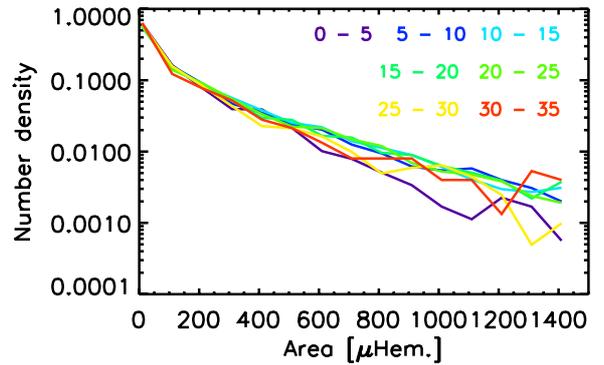
The differences between the two fits are mainly in the tails, which is a partial explanation of why both log-normal and power-law distributions have been reported in the past (see, e.g. Bogdan et al. 1988; Harvey & Zwaan 1993; Baumann & Solanki 2005; Harvey & Zwaan 1993; Schrijver & Harvey 1994). We also comment that we are here considering the sunspot group areas from the RGO record, which includes both the umbral and penumbral area but excludes the area of the plage. This also partially explains why our results can differ from those of previous authors.

#### 3.4.2. Cycle phase dependence of area distribution

We next consider dependence on the cycle phase and latitude of the area distribution. For the phase-of-cycle dependence we use the same type of analysis as in Sect. 3.1. Figure 9 shows the



**Fig. 9.** Mean sunspot group area for cycles 12–20. The red curve is the fit given by Eq. (14).



**Fig. 10.** Number density function for 5 degree-binned latitudes.

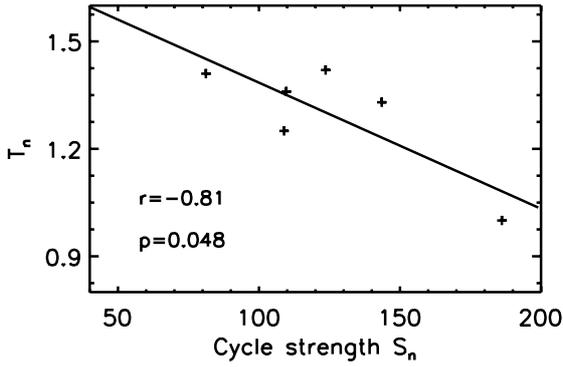
area distribution for different cycles as a function of the phase  $1 \leq i \leq 30$ . The average value over all cycles can be fitted by the second degree polynomial

$$A_i = 115 + 396(i/30) - 426(i/30)^2. \quad (14)$$

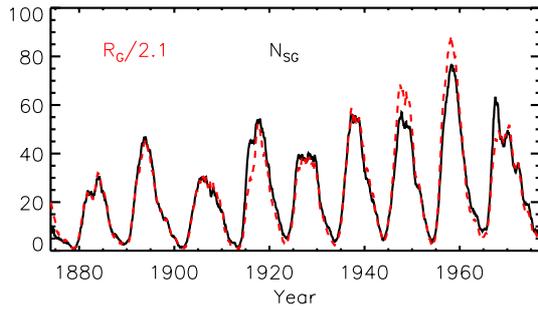
There is also a (possibly related) dependence of the areas on latitude. Figure 10 shows the number density function of sunspot group areas for 5 degree binned latitudes. Even after averaging the data in this way there is still some scatter apparent in the data. The relative scatter can, which can be judged from the latitude-to-latitude variation in the plot, increases with area for each latitude bin. This is because there are relatively few large sunspot groups. The figure also indicates that large sunspots rarely occur at low ( $<5$ ) latitudes. This partly reflects the phase dependence of the area distribution.

### 3.5. Tilt angle distribution

Joy's law (Hale et al. 1919) states that the line joining the centres of the positive and negative polarities of sunspot groups is systematically tilted with respect to the East-West direction. Recently, Dasi-Espuig et al. (2010) found that the tilt angles of sunspot groups, obtained from the Mount Wilson Observatory and from Kodaikanal observations, show a cycle-to-cycle variation. They further showed that the average tilt angle is negatively correlated with the strength of the cycle, i.e. the tilt angles tend to be smaller for stronger cycles. Incorporating the cycle-dependent tilt angles of sunspot groups in a surface flux transport model, Cameron et al. (2010) reproduced the empirically derived time evolution of the solar open magnetic flux and the reversal times of the polar fields from 1913 to 1986 (cycles 15–21), here we consider only the time period covered by the RGO data and



**Fig. 11.** The relationship between cycle strength  $S_n$  and  $T_n$ .



**Fig. 12.** A comparison of the number of sunspot groups appearing each month in the RGO data,  $N_{SG}$ , (black curve) and fit based upon  $R_G$  (red dashed curve).

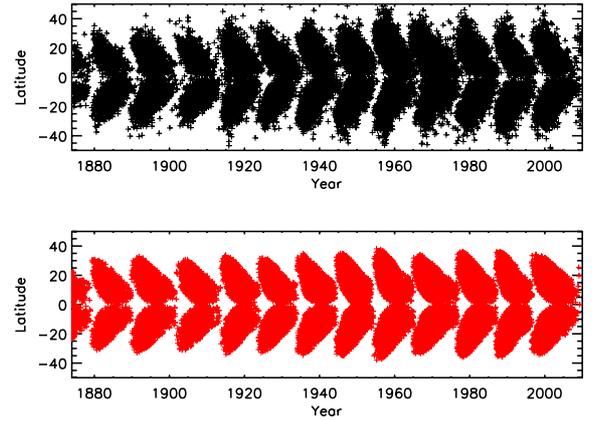
hence omit cycle 21 from our analysis. The precise dependence of the tilt angle on latitude is uncertain and it seems that a square root profile matches the data somewhat better than the usually assumed linear relationship. In this paper, as in CJSS10, we consider the square root profile  $\alpha_n = T_n \sqrt{|\lambda|}$ , where  $\alpha_n$  is the average tilt angle and  $T_n$  is the constant of proportionality for cycle  $n$ .

For each cycle we determined  $T_n$  in a similar way as in Dasi-Espuig et al. (2010) and Cameron et al. (2010). Figure 11 shows the correlation between cycle strength  $S_n$  and  $T_n$ . The correlation coefficient for  $S_n$  is 0.81, slightly higher than that of  $\tilde{S}_n$  which is 0.78. Under either definition, the cycle strength and  $T_n$  are significantly correlated ( $p < 0.05$ ). The linear fit between  $S_n$  and  $T_n$  is

$$T_n = 1.73 - 0.0035S_n. \quad (15)$$

### 3.6. Number of sunspot groups as a function of time

Our aim in this paper is to construct semi-synthetic sunspot group records, based on  $R_G$  and  $R_Z$ , with similar statistics to those of the RGO data. We have used  $R_G$  to determine strength of each cycle and have found correlations which allow us to construct synthetic latitudes, longitudes, areas and tilt angles for each spot group. We here determine how many sunspot groups should appear each month to make the semi-synthetic record have similar statistics as the RGO dataset. For the period covered by the RGO records, the monthly number should be approximately the same as the number of groups in the RGO dataset,  $N_{SG}$ . We have found that the fit  $R_G/2.1$ , shown in Fig. 12, matches the data well. We use this fit to reconstruct the number of sunspot groups appearing each month from 1700 onwards.



**Fig. 13.** A comparison of observed (upper panel) and semi-synthetic (lower panel) butterfly diagrams for the years 1874–2010.

## 4. Reconstruction of sunspot group emergence

Our goal is to reconstruct sunspot group emergence back to 1700 based solely on sunspot numbers. For that we use the derived relationships between sunspot groups (mean latitude, latitude width, longitude distribution, area distribution and tilt angles) and activity indices (sunspot numbers, strength, and phase of the cycles) to estimate the spatial and temporal properties of emerging sunspot groups. Since the relationships we have derived are only in terms of correlations, our models have a random component. For the latitudes and areas we draw from random populations which have the relevant distributions set out in the previous sections.

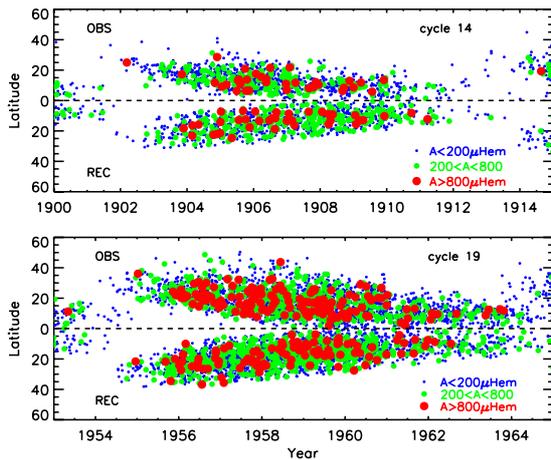
Implicit in such a reconstruction is the assumption that the dynamo has operated in a similar way from 1700 onwards. The very limited records of observations during the earlier part of the 18th century indicate that some of the early cycles might be anomalous in having stronger activity near the equator than those of the better observed later cycles (e.g., see Ribes & Nesme-Ribes 1993; Arlt 2009). This could indicate that the dynamo was operating in a not purely dipole mode during this period.

### 4.1. A semi-synthetic butterfly diagram covering 1874–2010

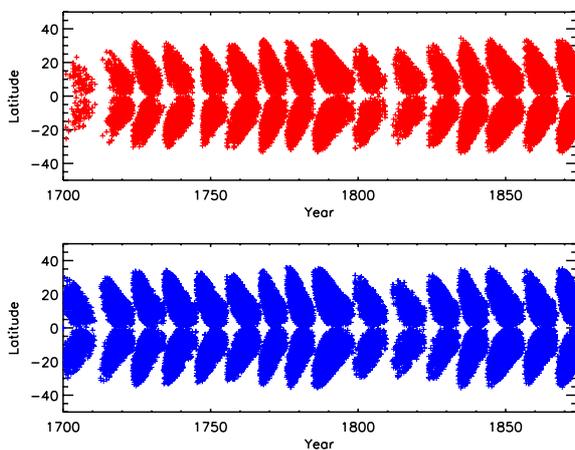
We first present an example semi-synthetic butterfly diagram for the period from the start of the RGO records to 2010. This allows us to directly compare the semi-synthetic and observed butterfly diagrams in Fig. 13. As expected, the two diagrams have similar appearances. A more detailed comparison of the weakest and strongest cycles is shown in Fig. 14. Again the observed and semi-synthetic butterfly wings look similar. This validates the use of the semi-synthetic reconstruction for periods when we only have the sunspot numbers.

### 4.2. Comparison of reconstructions using $R_G$ and $R_Z$ during 1700–1874

The semi-synthetic model shown in Figs. 13 and 14 was based on the group sunspot number  $R_G$ . Prior to 1874  $R_Z$  and  $R_G$  have substantial differences which affect the reconstructed butterfly diagrams. Figure 15 shows the reconstructed butterfly diagram during 1700–1874 with  $R_G$  and  $R_Z$ , respectively. It will be very interesting to compare both semi-synthetic butterfly diagrams with those being obtained by Arlt & Abdolvand (2010). We comment that there is no reason emerging from this study to prefer one data set over the other.



**Fig. 14.** Comparison of butterfly diagrams from observation (above the horizontal dashed lines) and reconstruction (below the dashed lines) for the weakest cycle 14 (upper panel) and the strongest cycle 19 (lower panel), both for the northern hemisphere. The area of the sunspot groups is indicated by the colors and sizes of circles.



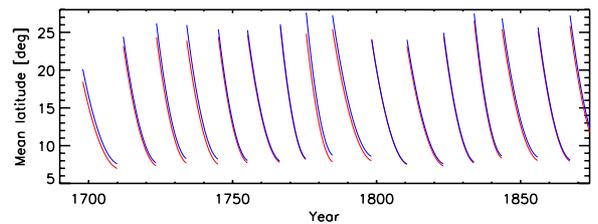
**Fig. 15.** Semi-synthetic butterfly diagram for the years 1700–1874 using  $R_G$  (upper panel) and  $R_Z$  (lower panel).

To give another indication of the differences in the reconstructions based on  $R_G$  and  $R_Z$  Fig. 16 shows the reconstructed mean latitudes during 1700–1874. The different cycle strengths derived from the two sets of sunspot numbers produce small differences which differ in strength from cycle to cycle. The extent to which these differences affect the results of surface flux transport simulations and the open flux calculated therefrom will be investigated in Paper II.

## 5. Conclusions

Using the group sunspot number  $R_G$  and RGO, MWO and Kodaikanal data sets, we studied the phase dependence and cycle dependence of latitude, area and tilt angle distribution properties of sunspot group emergence. The main correlations found are:

1. The mean latitude at which sunspots emerge can be modeled using a second order polynomial of cycle phase.
2. Strong cycles have a higher mean latitude for sunspot emergence (Fig. 3).
3. The ratio of the latitudinal range over which sunspot groups emerge and the average latitude of emergence varies as a function of cycle phase (Fig. 5).



**Fig. 16.** Mean sunspot latitudes for the years 1700–1874 reconstructed with  $R_Z$  (blue curves) and  $R_G$  (red curves).

4. The distribution of sunspot areas is similar for all cycles (Fig. 8).
5. The size distribution is a power-law for small sunspots and obeys a log-normal profile for large sunspots (Fig. 8).
6. During cycle maxima sunspots are, in the mean, larger (Fig. 9).
7. Sunspot nests are important, especially during cycle maximum phases (Fig. 6).

We have modeled and used the correlations to construct semi-synthetic butterfly diagrams extending back to 1700. This reconstruction will be useful in modeling the large-scale solar magnetic field over this period.

## References

- Arlt, R. 2009, *Sol. Phys.*, 255, 143  
 Arlt, R., & Abdolvand, A. 2010, [arXiv:1010.3131]  
 Bai, T. 1988, *ApJ*, 328, 860  
 Baumann, I., & Solanki, S. K. 2005, *A&A*, 443, 1061  
 Baumann, I., Schmitt, D., Schüssler, M., & Solanki, S. K. 2004, *A&A*, 426, 1075  
 Berdyugina, S. V., & Usoskin, I. G. 2003, *A&A*, 405, 1121  
 Bogdan, T. J., Gilman, P. A., Lerche, I., & Howard, R. 1988, *ApJ*, 327, 451  
 Bumba, V., & Howard, R. 1965, *ApJ*, 141, 1502  
 Cameron, R. H., Jiang, J., Schmitt, D., & Schüssler, M. 2010, *ApJ*, 719, 264  
 Castenmiller, M. J. M., Zwaan, C., & van der Zalm, E. B. J. 1986, *Sol. Phys.*, 105, 237  
 Chapman, G. A., Cookson, A. M., & Dobias, J. J. 1997, *ApJ*, 482, 541  
 Charbonneau, P. 2010, *Liv. Rev. Sol. Phys.*, 7  
 Crouch, A. D., Charbonneau, P., Beaubien, G., & Paquin-Ricard, D. 2008, *ApJ*, 677, 723  
 Dasi-Espuig, M., Solanki, S. K., Krivova, N. A., Cameron, R., & Peñuela, T. 2010, *A&A*, 518, A7  
 Devore, C. R., Boris, J. P., Young, Jr., T. R., Sheeley, N. R., & Harvey, K. L. 1985, *Australian J. Phys.*, 38, 999  
 Hale, G. E., Ellerman, F., Nicholson, S. B., & Joy, A. H. 1919, *ApJ*, 49, 153  
 Harvey, K. L., & Zwaan, C. 1993, *Sol. Phys.*, 148, 85  
 Harvey, K. L., & White, O. R. 1999, *J. Geophys. Res.*, 104, 19759  
 Hathaway, D. H. 2010, *Liv. Rev. Sol. Phys.*, 7, 1  
 Howard, R., Gilman, P. I., & Gilman, P. A. 1984, *ApJ*, 283, 373  
 Howard, R. F., Gupta, S. S., & Sivaraman, K. R. 1999, *Sol. Phys.*, 186, 25  
 Hoyt, D. V., & Schatten, K. H. 1998, *Sol. Phys.*, 179, 189  
 Ivanov, V. G., & Miletsky, E. V. 2011, *Sol. Phys.*, 268, 231  
 Li, K. J., Wang, J. X., Zhan, L. S., et al. 2003, *Sol. Phys.*, 215, 99  
 Ribes, J. C., & Nesme-Ribes, E. 1993, *A&A*, 276, 549  
 Schrijver, C. J., & Harvey, K. L. 1994, *Sol. Phys.*, 150, 1  
 Schrijver, C. J., & Zwaan, C. 2000, *Solar and Stellar Magnetic Activity*, Cambridge Astrophysics Series, 34 (Cambridge, UK: Cambridge University Press)  
 Sheeley, Jr., N. R., Devore, C. R., & Boris, J. P. 1985, *Sol. Phys.*, 98, 219  
 Sivaraman, K. R., Gupta, S. S., & Howard, R. F. 1993, *Sol. Phys.*, 146, 27  
 Solanki, S. K., Wenzler, T., & Schmitt, D. 2008, *A&A*, 483, 623  
 Usoskin, I. G. 2008, *Liv. Rev. Sol. Phys.*, 5, 3  
 van Ballegoijen, A. A., Cartledge, N. P., & Priest, E. R. 1998, *ApJ*, 501, 866  
 Wang, Y., Nash, A. G., & Sheeley, Jr., N. R. 1989, *Science*, 245, 712  
 Wolf, R. 1861, *MNRAS*, 21, 77  
 Zhang, J., Wang, Y., & Liu, Y. 2010, *ApJ*, 723, 1006  
 Zhang, L. Y., Wang, H. N., Du, Z. L., Cui, Y. M., & He, H. 2007, *A&A*, 471, 711  
 Zhao, X., & Hoeksema, J. T. 1995, *Space Sci. Rev.*, 72, 189