Active Longitude and Coronal Mass Ejection Occurrences

N. Gyenge\textsuperscript{1,2}, T. Singh\textsuperscript{3}, T. S. Kiss\textsuperscript{1,4}, A. K. Srivastava\textsuperscript{3}, and R. Erdélyi\textsuperscript{1}

\textsuperscript{1} Solar Physics and Space Plasmas Research Centre (SP2RC), School of Mathematics and Statistics, University of Sheffield Hounsfield Road, Hicks Building, Sheffield S3 7RH, UK; n.g.gyenge@sheffield.ac.uk
\textsuperscript{2} Debrecen Heliophysical Observatory (DHO), Konkoly Observatory, Research Centre for Astronomy and Earth Sciences Hungarian Academy of Sciences, Debrecen, P.O. Box 30, H-4010, Hungary
\textsuperscript{3} Department of Physics, Indian Institute of Technology (Banaras Hindu University), Varanasi, India
\textsuperscript{4} Department of Physics, University of Debrecen, Egyetem tér 1, Debrecen, H-4010, Hungary

Received 2016 October 24; revised 2017 February 21; accepted 2017 February 21; published 2017 March 20

Abstract

The spatial inhomogeneity of the distribution of coronal mass ejection (CME) occurrences in the solar atmosphere could provide a tool to estimate the longitudinal position of the most probable CME-capable active regions in the Sun. The anomaly in the longitudinal distribution of active regions themselves is often referred to as active longitude (AL). In order to reveal the connection between the AL and CME spatial occurrences, here we investigate the morphological properties of active regions. The first morphological property studied is the separateness parameter, which is able to characterize the probability of the occurrence of an energetic event, such as a solar flare or CME. The second morphological property is the sunspot tilt angle. The tilt angle of sunspot groups allows us to estimate the helicity of active regions. The increased helicity leads to a more complex buildup of the magnetic structure and also can cause CME eruption. We found that the most complex active regions appear near the AL and that the AL itself is associated with the most tilted active regions. Therefore, the number of CME occurrences is higher within the AL. The origin of the fast CMEs is also found to be associated with this region. We concluded that the source of the most probably CME-capable active regions is at the AL. By applying this method, we can potentially forecast a flare and/or CME source several Carrington rotations in advance. This finding also provides new information for solar dynamo modeling.

Key words: Sun: activity – Sun: coronal mass ejections (CMEs) – Sun: flares – Sun: magnetic fields – sunspots

1. Solar Nonaxisymmetric Activity

Since the beginning of the previous century, the nonhomogeneous spatial properties of solar activity have been studied extensively (Carrington 1863; Maunder 1905; Chidambaram 1932; Losh 1939; Bumba & Howard 1965; Bumba & Obridko 1969). These early investigations conjectured initially that the longitudinal distribution of sunspot groups or sunspot numbers shows nonhomogeneous behavior. These analyses concluded that there are preferred longitudes, where solar activity concentrates.

Later, different approximations and assumptions were applied to understand the essence of this phenomenon. The topic soon became controversial (see, e.g., Henney & Durney 2005; Pelt et al. 2006). Overall, three approaches can be distinguished. The first approach is the quasi-rigid structure model by Warwick (1966). This model describes a constantly rotating frame that carries the persistent domains of activity (Ivanov 2007). Bogart (1982) applied an autocorrelation statistical method based on long-term sunspot number data. Balthasar & Schüssler (1983, 1984) applied period analysis to the Greenwich Photoheliographic Results. These studies concluded that the angular velocity of the quasi-rigid rotating frame varies. The angular velocity depends on the solar cycle, but during one cycle the angular velocity is constant.

The second approach, promoted by, e.g., Becker (1955), Castenmiller et al. (1986), and Brouwer & Zwaan (1990), discovered the “active nest” and defined it as a small and isolated area on the solar surface. Here, the enhanced longitudinal activities are considered as individual entities. These isolated entities can be absent for several rotations.

The third group of models assumes a migrating activity in the Carrington coordinate system. Berdyugina & Usoskin (2003) and Berdyugina (2004, 2005) found persistent active longitudes (ALs) under the influence of the differential rotation. Usoskin et al. (2005) and Berdyugina et al. (2006) introduced a “dynamic reference frame.” This frame describes the longitudinal migration of AL in the Carrington coordinate system and has a similar dynamics to differential rotation. Usoskin et al. (2007) concluded that the migration of the enhanced activity is just apparent. In the “dynamic reference frame,” the rotation of the AL remains constant and the AL itself is a persistent quasi-rigidly rotating phenomenon. Usoskin et al. (2007) proposed that a seemingly migrating AL may occur as a result of interaction between the equatorward-propagating dynamo wave and a quasi-rigidly rotating nonaxisymmetric active zone. The role of differential rotation is also debated. Various studies concluded that the differential rotation is not the reason for the migration of the AL (Balthasar 2007; Juckett 2007; Gyenge et al. 2014).

In our previous work (i.e., Gyenge et al. 2016, hereafter GY16), we found evidence supporting the third group of models. Moreover, the migration of ALs does not appear to correspond very well to the form of the 11 yr cycle as suggested by previous studies (Usoskin et al. 2005; Berdyugina et al. 2006). The half-width of the active longitudinal belt is fairly narrow during moderate activity but is wider at maximum activity. We also found that the AL is not always identifiable.

Several studies (Warwick 1965; Bai 2003a, 2003b) suggested that the spatial distributions of eruptive solar phenomena also show nonaxisymmetric properties. Bai (1987, 1988) analyzed the coordinates of energetic solar flares based on a 5 yr time period and concluded that longitudinal spatial distribution is nonhomogeneous. Zhang et al. (2007, 2008, 2015) concluded that the dominant and co-dominant AL
contains 80% of C- and X-flares. In GY16, we conducted a similar study based on four solar cycles, and we did not find significant co-dominant activity; instead, we found that only the dominant AL contains 60% of the solar flares.

The flares and coronal mass ejections (CMEs) could occur independently of each other. Numerous CMEs have associated flares, but several nonflaring filament lift-offs also lead to CME (Gosling et al. 1976; Harrison 1995). Furthermore, in the case of the “stealth” CMEs there are no easily identifiable signatures to locate the source of the eruption on the solar surface (Howard & Harrison 2013). Hence, separate flare—CME spatial distribution investigations are justified.

2. CME and Sunspot Data

We used the SOHO/LASCO HALO CME catalog by the CDAW data center. The CME catalog spans over 20 yr, i.e., between 1996 and 2016. This is the most extensive catalog that contains the source location of CMEs. Only halo CMEs are reported, i.e., their angular width is 360° in the C3 coronagraph field of view. Gopalswamy et al. (2009) describe the catalog in great detail. The CME source is defined as the center of the associated active region (AR). The source of the CME is identified using SOHO EIT running difference images. Later, the ability of the STEREO mission to observe the backside of the Sun was used to identify the source of back-sided halo CMEs (Gopalswamy et al. 2015). This catalog also provides the speed of CMEs, which is the actual speed with which the CME propagates in the interplanetary space. The plane-of-sky speed is obtained from the single SOHO viewpoint, converted into space speed using the cone model (Xie et al. 2004).

The Debrecen Photoheliographic Data (DPD) sunspot catalog and sunspot tilt angle catalog are used for estimating the longitudinal position of the AL (Győri et al. 2017; Baranyi 2015). The catalog provides information about the date of observation, position, and area for each sunspot. The precision of the position is 0.1 heliographic degrees, and the estimated accuracy of the area measurement is ~10%.

3. Longitudinal Sunspot Distribution

The first step of our identification procedure is to divide the solar surface by 18 equally sliced longitudinal belts. Hence, one bin is equal to a zone with 20° width. We take into account all sunspot groups from the moment when they reach their maximum area. This filtering criterion is chosen for the following reasons. First, if we select all sunspot groups at every moment of time, then the statistics will be biased by the long-lived sunspot groups. Second, the maximum area of the sunspot groups is a well-defined and easily identifiable moment. This is different from the used practice of considering only the first appearance of each sunspot group. Let us define the matrix \( W \) by

\[
W_{\lambda,\text{CR}} = \frac{A_{\lambda,\text{CR}}}{\sum_{i=1}^{n} A_{i,\text{CR}}}. \tag{1}
\]

Here, the area of all sunspot groups \( A_{\lambda,\text{CR}} \) is summed up in each longitudinal bin for each Carrington rotation (CR). \( A_{i,\text{CR}} \) is divided by the summarized area over the entire solar surface \( \sum_{i=1}^{n} A_{i,\text{CR}} \). The range of \( W \) must be always between 0 and 1, depending on the local appearance of the activity. In the case of \( W_{\text{CR}} = 1 \), all of the flux emergence takes place in one single longitudinal strip. A 3\( \sigma \) significance level threshold is applied to filter the noise. Moving average is also applied for data smoothing with a time window of three CRs.

We standardized the matrix \( W_{\lambda,\text{CR}} \) (defined in Equation (1)) by removing the mean of the data and scaling to unit variance. Then, cluster analysis was performed for grouping the obtained significant peaks. Here, the DBSCAN clustering algorithm was chosen, which is a density-based algorithm. The method groups together points that are relatively closely packed together in a high-density region, and it marks outlier points that stand alone in low-density regions (for details see Ester et al. 1996). The parameter epsilon (=0.2) defines the maximum distance between two points to be considered to be in the same group. The parameter \( m (=3) \) specifies the desired minimum cluster size. Clusters containing fewer than three points were omitted. The longitudinal location of the clusters \( \lambda_{\text{cluster,CR}} \) represents the position of the AL.

Panel (a) of Figure 1 shows an example of the initial identification steps outlined above. The sample time period covers 6 yr, and it corresponds to 80 CR between 1920 CR and 2000 CR. The quantity \( W \) is represented by the different shades of blue. The dark-blue regions denote the significant (3\( \sigma \)) presence of activity. The lighter shades stand for a weaker manifestation of activity (2\( \sigma \) and 1\( \sigma \)). The gray squares shows the sunspot group clusters.

In panel (b) of Figure 1, the Carrington longitudes of the most significant cluster are transformed into Carrington phase for each CR:

\[
\psi_{\lambda,\text{CR}} = \frac{\lambda_{\text{cluster,CR}}}{360}, \tag{2}
\]

The range of the quantity \( \psi_{\lambda} \) must be between 0 and 1, where \( \psi_{\lambda} = 1 \) represents the entire circumference of the Sun. In panel (b) of Figure 1, panel (a) is repeated three times so that we are able to track the migration of the activity throughout the phases. To the data panels (a) and (b), we applied a polynomial least-squares fitting based on multiple models. Linear, quadratic, cubic, and higher-order polynomial models were tested. The quadratic or parabolic regression \( (Ax^2 + Bx - C) \) shows the best goodness of fit; hence, this model is chosen. Table 1 shows the coefficients and uncertainties.

The shape of migration clearly follows a parabolic-shaped path as found in several earlier studies (Berdyugina & Usoskin 2003; Berdyugina et al. 2006; Usoskin et al. 2007, 2005; Zhang et al. 2011a; Gyenge et al. 2014).

In panel (c) of Figure 1, we plot the parabola-shaped migration path in a coordinate system that follows the parabolic shape of panel (b). Hence, the AL is easily recognizable without repeating Carrington phases. Several new features are noticeable in panel (c): parallel aligned and tilted lanes between CR 1930 and CR 1950. These shapes are artificial, created by the coordinate system transformation. In the new dynamic coordinate system the AL stands still as represented by the black regression line.

Panels (d), (e), and (f) of Figure 1 show the kernel density estimation (KDE) of the spatial difference between the AL \( \psi_{\lambda,\text{AL}} \) and the longitudinal position of individual sunspot groups \( \psi_{\text{CR}} \) in Carrington phase difference \( \Delta \psi \); see Equation (3) applied to the entire period investigated (i.e., between 1997 January 06 and 2015 December 30) and for both

---

3. Longitudinal Sunspot Distribution

3.1. The longitudinal migration of the AL

3.2. Relationship between activity and migration

3.3. The long-term variability of the AL

---

http://cdaw.gsfc.nasa.gov/CME_list/halo/halo.html
http://fenyi.solarobs.unideb.hu/DPD/
hemispheres:

$$\Delta \psi^* = |\psi_{AL} - \psi_{CR}|.$$  

$$\Delta \psi = \begin{cases} 
\Delta \psi^*, & \text{if } \Delta \psi^* \leq 0.5, \\
1 - \Delta \psi^*, & \text{if } \Delta \psi^* > 0.5.
\end{cases}$$  \hspace{1cm} (3)

The KDE is a nonparametric method to estimate the probability density function (Connolly et al. 2000). We used a Gaussian kernel function. The optimal value of the Gaussian kernel bandwidth is 0.02. If \(\Delta \psi = 0\), the sunspot group is located at the AL. In the case of \(\Delta \psi = 0.5\), the locations of the sunspot groups are shifted by 180° from the position of the AL. This shifted position is marking the area of the co-dominant AL. The three KDEs are obtained from different sunspot group data sets with different threshold levels. Panel (f) of Figure 1 shows the sunspot distribution around the AL with different significance levels taken for the entire time period and for both hemispheres. The horizontal axes \(\Delta \psi\) represent the shortest distance between the migration of enhanced longitudinal activity and given sunspot groups. Panel (g) is the cumulative distribution of the above three probability density functions (PDFs).

**Figure 1.** Panels (a), (b), and (c) show an example of the migration of the AL between 1920 CR and 2000 CR (1997 March 01–2003 February 20) based on data of the solar northern hemisphere. The shades of blue indicate the significance of sunspot group activity. The gray squares show high-density areas, i.e., the detected clusters. The solid black line represents the migration patch, fitted by the least-squares method and considering only the most significant (above 3σ) clusters. Panel (a) shows the observed longitudinal distribution of sunspots in the Carrington coordinate system. Panel (b), similar to panel (a), depicts solar circumference (Carrington phase) repeated three times. Panel (c) is the phase-corrected migration path. The parabolic migration pattern is now transformed to a constant line and provides an insight into the nonhomogeneous spatial property. Panels (a) and (c) use the same color scale as defined in panel (b). The color scale is displayed in the lower right corner of panel (b). Panels (d), (e), and (f) show the sunspot distribution around the AL (corresponding to \(\Delta \psi = 0\)) with different significance levels taken for the entire time period and for both hemispheres. The horizontal axes \(\Delta \psi\) represent the shortest distance between the migration of enhanced longitudinal activity and given sunspot groups.
more disperse distribution. Below the 1σ threshold, the KDE does not show obvious peaks. Panel (g) of Figure 1 depicts cumulative distribution functions (CDFs) of the three KDEs. Note that 70% of the most significant sunspot groups appear closer than $0.17$ $\Delta \psi = 0.17$. This value corresponds to a longitudinal zone with $\pm 60^\circ$ of width around AL. Figure 2 shows the migration path of the AL based on the entire analyzed period. The coefficients and uncertainties of the employed quadratic model functions are given in Table 1.

### Table 1

| Solar Cycle | Hemisphere | Coefficient A | Coefficient B | Coefficient C |
|-------------|------------|---------------|---------------|---------------|
| 23 North    | $-4.32 \times 10^{-4} \pm 5.44 \times 10^{-5}$ | $1.71 \pm 2.15 \times 10^{-1}$ | $1.69 \times 10^{1} \pm 2.13 \times 10^{2}$ |
| 23 North    | $-5.61 \times 10^{-4} \pm 1.78 \times 10^{-4}$ | $2.37 \pm 7.47 \times 10^{-1}$ | $2.51 \times 10^{0} \pm 7.83 \times 10^{2}$ |
| 23 South    | $-5.62 \times 10^{-4} \pm 6.62 \times 10^{-5}$ | $2.23 \pm 2.62 \times 10^{-1}$ | $2.20 \times 10^{0} \pm 2.59 \times 10^{2}$ |
| 24 South    | $-5.16 \times 10^{-4} \pm 1.85 \times 10^{-5}$ | $2.18 \pm 7.77 \times 10^{-1}$ | $2.31 \times 10^{0} \pm 8.12 \times 10^{2}$ |

### 4. Complexity Properties of AL Sunspots

#### 4.1. Separateness Parameter

The relationship between flare occurrences and the morphological properties of sunspot groups is widely accepted and is reported in numerous studies (e.g., Cui et al. 2006; Schrijver 2007; Mason & Hoeksema 2010; Korsós et al. 2014, 2015). The first classification scheme (Hale et al. 1919) is published by Waldmeier (1938). The scheme examines the role of the size and morphology of sunspot groups in relation to determining the capacity of their flare productivity. The system was modified by Waldmeier (1947) and is known today as the modified Zürich classification system (Kiepenheuer 1953). The classification was further developed by McIntosh (1990), and this version is still in use today. Later the classification was automated by Colak & Qahwaji (2008). However, the classification of the Zürich or McIntosh system is still subjective; further, there are just a limited number of classes (Bornmann & Shaw 1994).

The separateness parameter ($S$) is used to reveal the morphological properties of the sunspot groups near and far from the AL. This parameter was introduced by Korsós &
Erdélyi (2016), and its investigation and application to flares showed that the separateness parameter can be a numerical indicator/precursor besides the traditional (Zürich, McIntosh, and Mount Wilson) classifications of sunspots. The parameter $S$ is considered as an indicator for the potential flaring outbreak and CME capability of ARs.

The separateness parameter is determined by the angular distance between the area-weighted centers of the leading and following subgroups divided by the angular diameter of a hypothetical circle whose area is equal to the total area of all umbrae constituting the sunspot group.

The angular distance is the shortest distance between two points on the surface of a sphere. The distance (in degrees) between the leading and following subgroups is provided by the spherical law of cosines:

$$
\Delta \theta = 2 \arcsin \left[ \sin^2 \left( \frac{|B_l - B_f|}{2} \right) + \cos(B_l) \cos(B_f) \sin^2 \left( \frac{|L_l - L_f|}{2} \right) \right].
$$

(4)

Here, $B$ and $L$ refer to the heliographical latitude and longitude of $l$ leading and $f$ following subgroups. If the absolute difference is greater than 180 ($|L_l - L_f| > 180$), then the absolute difference is $360 - |L_l - L_f|$. The corrected area of individual sunspots ($A^*$) in millionths of solar hemisphere is converted to Mm$^2$, using

$$
A = \frac{1}{2} \left( 4\pi R_{\text{Sun}}^2 \right) 10^{-7} A^*,
$$

(5)

where $R_{\text{Sun}}$ is in Mm. The total sunspot group area ($T$) is calculated. The number of sunspots in a certain sunspot group is represented by the quantity $n$. The total area (Mm$^2$) means the summed-up area of the individual sunspots:

$$
T = \sum_{i=1}^{n} A_i.
$$

(6)

The diameter of an individual sunspot group, ($\Delta \Omega$), is estimated by

$$
\Delta \Omega = \frac{2 \sqrt{T/\pi}}{2 R_{\text{Sun}} \cos \left( \frac{1}{2} (B_l + B_f) \pi \right)} \times 360^\circ.
$$

(7)

The numerator is the diameter (Mm) of a hypothetical circle whose area is equal to the total area $T$. The denominator represents the circumference of a small circle (Mm) that connects all locations with a given latitude. The fraction is multiplied by $360^\circ$, which is equal to the angular distance between the endpoints of the sunspot group diameter in degrees.

Finally, let us define the dimensionless separateness parameter:

$$
S = \frac{\Delta \theta}{\Delta \Omega}.
$$

(8)

In Figure 3, typical ARs NOAA AR 11429, 11666, and 11241 (from top to bottom) are selected to demonstrate the usefulness of the parameter $S$. The visualization of the ARs is plotted on the left-hand side. The blue and red colors distinguish the different magnetic polarities, and the radius of a circle represents the area of the spot. The black dots indicate the weighted average position of the leading and following subgroups. The black dashed line between the black dots is the calculated angular distance ($\Delta \theta$), described by Equation (4). The gray circle around the spots is the hypothetical circle whose area is equal to the total area of all umbrae constituting the sunspot group ($\Delta \Omega$), defined by Equation (7). Panels on the right-hand side are the snapshots (HMI magnetogram by SDO) of the ARs.

The upper two panels of Figure 3 are visualizing a complex sunspot group (namely, NOAA AR 11429). This sunspot group is a beta-gamma-delta magnetic configuration according to the Mount Wilson classification. The AR is extremely complex, having umbrae of opposite polarity within the same penumbra. The calculated separateness parameter is 0.6 ± 0.25 (Equation (8)). The middle panels of Figure 3 show NOAA AR 11666, which is a moderate complex sunspot group ($S = 1.3 \pm 0.35$). The bottom panels display NOAA AR 11241, a less complex bipolar sunspot group ($S = 2.0 \pm 0.30$).

Based on the study by Korsós & Erdélyi (2016), there is a high risk of X-class flare and/or fast CME occurrence(s) if $S < 1$. There is a moderate risk of flaring (M class) and/or CME occurrence(s) if $S > 1$ and $S < 2$. In the case of $S > 3$, only bipolar sunspot groups appear with relatively simple morphological properties. These ARs have a rather low probability of a significant flaring (above GOES C class) or CME activities.

### 4.2. Separateness Parameter within AL

In this section, we investigate the longitudinal spatial distribution of the parameter $S$ (Equation (8)) and the area of the investigated sunspot groups. The spatial distance of sunspot groups ($\Delta \psi$) is defined by Equation (3).

The raw area measures are converted to standard scores (standardized statistics). The standard score is a dimensionless quantity calculated by subtracting the average of the sunspot group area ($\bar{T}$) from a given group area $T_i$ and dividing by the sample-corrected standard deviation ($\sigma(T)$) of the area data:

$$
Z_i = \frac{T_i - \bar{T}}{\sigma(T)}.
$$

(9)

Spatial multivariable (linear) interpolation is applied to $f(S, \Delta \psi, Z)$. The method results in a regular matrix (M) from unstructured three-dimensional data. The range of $\log S$ is $[-1, 0.5]$ and is divided by 1500 equal bins ($n$). The range of $\Delta \psi [0, 0.5]$ is divided by 500 bins ($m$), i.e.,

$$
M = \begin{bmatrix}
Z_{1,1} & Z_{1,2} & \cdots & Z_{1,n} \\
Z_{2,1} & Z_{2,2} & \cdots & Z_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
Z_{m,1} & Z_{m,2} & \cdots & Z_{m,n}
\end{bmatrix}.
$$

(10)

Figure 4 shows the results of the statistics. Panels (a), (b), and (c) are obtained for data of the northern hemisphere (red plots), and panels (d), (e), and (f) are for those of the southern hemisphere (blue plots). In panels (a) (northern hemisphere) and (f) (southern hemisphere), the matrix $M$ (Equation (10)) is visualized. The horizontal axis is the distance from AL. The vertical axis is the logarithm of parameter $S$. $\log S < 0$ stands for a high-risk flare or CME occurrence. The color code is the standard score of the sunspot group area. The red or blue shades indicate positive standard score ($T_i > \bar{T}$). The white
color stands for negative standard score ($T < T_i$ or no data). Panels (b) and (e) are the row averages of matrix $M$. Panels (c) and (d) are the column averages of matrix $M$.

In panel (a), significant islands are visible between $0 > \log S > -0.5$ at $\Delta \psi < 0.1$. There is one more obviously visible island around $\Delta \psi = 0.2$ at $\log S < -0.4$. However, above $\Delta \psi > 0.2$ there is no remarkable island. Panel (c) of Figure 4 also reveals a peak below $\Delta \psi < 0.1$. The statistics suggest that the most complex and largest sunspot groups appear near the AL.

Analysis of the data of the southern hemisphere show similar results. There are easily notable islands below $\Delta \psi < 0.2$ at

Figure 3. Representation of the reconstructed sunspot groups (left-hand side) using the Debrecen sunspot Data (DPD) catalog applied to ARs NOAA AR 11429 (2012 March 07 15:02:44), NOAA AR 11666 (2011 March 01 15:02:45), and NOAA AR 11241 (2011 June 25 15:02:26). The polarities of the spots are distinguished by the different colors (red and blue). The gray color is the hypothetical circle, having summarized the area of the group. The dashed line is the distance between the following and leading subgroups. The area is measured in MSH (millionths of solar hemisphere). The sunspots are corrected for foreshortening. On the right-hand side, the magnetograms of the example sunspot groups are displayed.
\[
\log S = -0.6. \text{ Panels (d) and (e) of Figure 4 clearly reveal that significant sunspot groups appear only near the AL.}
\]

Both statistics suggest that the most complex ARs tend to cluster near the AL. The co-dominant AL around \(\Delta \psi = 0.5\) does not have significant activity. This statistical investigation also highlights a nonequivalent AL and co-dominant AL activity.

### 4.3. Tilt Angle of Investigated Active Regions within AL

The last investigated morphological property of ARs, here, is the sunspot group tilt angle. The definition of the tilt angle \(\gamma^*\) is given by Howard (1991):

\[
\gamma^* = \frac{(B_f - B_l)/(L_f - L_l)}{\text{sign}(|B_f| - |B_l|)\cos(B)}.
\]

Parameters \(B\) and \(L\) are the Carrington latitude and longitude, respectively. The following and leading subgroups have subscripts \(f\) and \(l\), respectively.

Figure 5 demonstrates the relationship found between the separateness parameter \(S\) and scaled tilt angle \(\gamma = |\gamma^*|/90\). Only the most significant ARs are taken into account; the area of sunspot groups has to be at least \(2\sigma\) greater than average.
The results applicable to the two hemispheres are distinguished by the colors of dots.

Linear regression cannot be used because the data are associated with a considerable uncertainty in both the $X$ and $Y$ directions (Isobe et al. 1990). For that reason, principal component analysis (PCA) is used to fit the data set. The PCA method is a linear dimensionality reduction keeping only the most significant singular vectors to project the data to a lower dimensional space (Einbeck et al. 2007). The eigenvector of the first component $e_1 = [-0.8387, 0.5445]$ shows the direction of the maximum variance of the data ($\sigma^2 = 4.1227$), i.e., where the data are most spread out.

Based on the result of the PCA, we performed principal component regression. In Figure 5, the solid black line is the regression fit and the gray halo represents a $1\sigma$ standard deviation of the sample along the regression line. The difference between the samples of two hemispheres is statistically insignificant; therefore, the regression was applied to the data of both hemispheres. The obtained statistics suggest that there is a clear relationship between the tilt angle and the separateness parameter of the sunspot groups.

5. Enhanced Longitudinal Behavior of CME Events

5.1. Spatial Possibly CME Occurrences

In this section, the connection between the CME occurrence and AL is revealed. Panel (a) of Figure 6 shows the kernel probability density function of the longitudinal distribution of CME occurrence. This statistic is based on data from both the northern and southern hemispheres. There is only one significant peak visible around $\Delta\psi = 0.05$. Besides this remarkable peak, there is a long plateau with some insignificant local peaks. Only one more peak, above the significance level at $\Delta\psi = 0.4$, is present, but with a relatively weak activity when compared to the first peak.

A random-generated control group is also used in this statistic. The longitudinal position of AL now is a random position. This test was inspired by Pelt et al. (2005), who expressed a critical view on the identification method of Berdyugina & Usoskin (2003) employed for AL. In our study, we applied the methodology introduced by Pelt et al. (2005), who reconstructed the distribution of AL with random sunspot longitude data. The KDE plot of the control group does not show any peaks. This homogeneous distribution means that AL identification does not cause false significant peaks, which would affect the results.

The lower panel of Figure 6 shows the cumulative distribution of the above-defined spatial distributions. The blue and red lines have a steep increasing phase between values of 0 and 0.1 followed by a less steep increasing trend. These results allow us to estimate that most CMEs (around 60%) occur in a $\pm 36^\circ$ belt around the position of AL. Hence, the width of the longitudinal belt of CME occurrences is equal to the width of the longitudinal belt of solar flare occurrences (GY16). The black line is the cumulative distribution obtained from the analysis applied to random longitudinal positions. This distribution would only contain 20% of CMEs. The latter...
finding means that AL plays a significant role in the spatial distribution of CME occurrences.

5.2. CME Dynamics

Let us now consider the apparent and space velocities ($V_a$ and $V_s$) of CME events. Two-dimensional kernel density estimations are applied with an axis-aligned bi-variate Gaussian kernel, evaluated on a square grid of the $\Delta \psi-V_a$ and $\Delta \psi-V_s$ space. Figure 7 shows the result, based on data from both hemispheres. The significance levels $1 - 3\sigma$ are indicated by colored contour lines.

In both panels of Figure 7, there are four islands above the $1\sigma$ significance level. The statistics show that the source of fast CMEs (speeds between 1500 and 3000 km s$^{-1}$) is indeed an AR, located within AL. However, slow (i.e., speed less than 1500 km s$^{-1}$) CMEs can occur outside of AL.

Above the significance level of $3\sigma$, there are only two islands. These are only slow CMEs inside and outside of AL. Analysis of this statistic also indicates that the probability of a slow CME is two standard deviation units higher than the probability of a fast CME.

6. Discussion

The AL identification method presented here reveals new spatial properties of the longitudinal distribution of the sunspot groups (panels (d), (e), and (f) of Figure 1). The spatial distribution of smaller sunspot groups (between $1\sigma$ and $2\sigma$) show already inhomogeneous properties. However, these results still have to be treated with caution. Only sunspot groups above the $3\sigma$ significance level have signatures of obvious and remarkable inhomogeneous spatial distribution.

The idea of two, almost equally significant longitudinal zones is widely accepted by numerous studies (see, e.g., Berdyugina & Usoskin 2003). The dominant and co-dominant AL is separated by 180$^\circ$ (Bumba et al. 2000; Zhang et al. 2011b). However, we do not find such equally strong ALs (neither here nor in GY16). In our investigation, the co-dominant AL plays a less important role.

The spatial distribution of the separateness parameter (defined by Equation (8)) shows that complex ARs with a high CME capability appear mostly near the AL (Figure 4). Moderate and simple complex configurations appear everywhere on the solar surface. These groups are also able to have CMEs with a significantly lower probability.

We also found that the most tilted sunspot groups have a complex configuration (Figure 5). Simple bipolar sunspot groups show relatively small tilt angle. Sakurai & Hagino (2003) and Canfield & Pevtsov (1998) concluded that there is positive correlation between magnetic helicity and sunspot tilt angle. The sunspot rotation could play an important role in helicity transport across the photosphere. Sunspot rotation may increase helicity in the corona, leading to flares and CMEs (Pevtsov 2012). This property may also have the consequence of a more complex buildup of the underlying magnetic structure, and the well-studied magnetic arches of the upper solar layer could be oriented at a large angle to the equator (Grigorev et al. 2012). The more complex ARs are, the more flares there are, and they will be associated with CMEs (Jetsu et al. 1997; Kitchatinov & Olemskoi 2005; Huang et al. 2013). Hence, we conclude that the above physical process can take place within AL and anywhere else, but only with low probability there.

Several studies have investigated east–west asymmetry of CME occurrence. Skrjigielo (2005) found asymmetry using data provided by the SOHO-LASCO in 1996–2004. The asymmetric behavior could be a consequence of AL. Our result obtained here shows (see, e.g., Figure 7) that the amount of CME occurrence is marginally higher within AL. The mean of the apparent and space velocity of CME occurrences is around 500 km s$^{-1}$ considering the entire surface of the Sun. This mean velocity is known as “slow” CMEs, found in a number of earlier studies (Ying et al. 2016). However, the mean velocity is significantly higher if only the AL itself is considered. Within AL the average (or space) velocity is around 1000 km s$^{-1}$ (see, e.g., Michalek et al. 2009). There is no fast CME occurrence found outside of AL. Therefore, interestingly and notably, the fast HALO CMEs are also AL CMEs.
7. Conclusion

Our new findings (together with the results of GY16) could provide novel aspects both for space weather forecast and for solar dynamo theory. Usually, the flare and/or CME prediction tools are based on only the behavior of ARs, such as complexity of magnetic fields or other morphological properties. However, the spatial distribution of ARs can also assist in forecasting, as suggested by, e.g., Zhang et al. (2008). We conclude that the main source of CME and solar flare (GY16) occurrences is the AL. Hence, the detection of this enhanced longitudinal belt may allow us to find the most flare- and CME-capable regions of the Sun preceding the appearance of an AR. This potential flare and CME source is predictable even several solar rotations in advance.

The observed properties of the nonaxisymmetric solar activity need to be taken into account in developing and verifying suitable dynamo theory: the observations analyzed here show that there is only one significant AL with a relatively wide ($\pm 20^\circ$–$30^\circ$) belt. Furthermore, the tilt angle of the ARs is also an important observed constraint for dynamo theory: the tilt angle of sunspot groups shows nonaxisymmetric behavior, which is a completely new (and surprising) finding.

The results of this research were enabled partially by SunPy, an open-source and free community-developed Python solar data analysis package (Mumford et al. 2015). N.G. is thankful for the support received from the University of Sheffield. N.G. is also thankful for the laborious work done by the assistant fellows at Debrecen Heliosophysical Observatory for composing the DPD sunspot catalog. R.E. acknowledges the support received by the Chinese Academy of Sciences Presidents International Fellowship Initiative, Grant No. 2016VMA045, the Science and Technology Facility Council (STFC), and the UK and Royal Society (UK). A.K.S. acknowledges the RESPOND-ISRO (DOS/PAOGIA205-16/130/602) project and the SERB-DST project (YSS/2015/00621) grant.