The Astrophysical Journal, 952:97 (10pp), 2023 August 1 https://doi.org/10.3847/1538-4357/acdb65

Statistical Study of the Correlation between Solar Energetic Particles and Properties of Active Regions

Russell D. Marroquin[1,2], Viacheslav Sadykov[2], Alexander Kosovichev[3,4], Irina N. Kitiashvili[4], Vincent Oria[5], Gelu M. Nita[3], Egor Illarionov[6,7], Patrick M. O'Keefe[5], Fraila Francis[5], Chun Jie Chong[5], Paul Kosovich[3], and Aatiya Ali[2]
[1] Department of Physics, University of California San Diego, La Jolla, CA 92093, USA
[2] Physics & Astronomy Department, Georgia State University, Atlanta, GA 30303, USA; vsadykov@gsu.edu
[3] Physics Department, New Jersey Institute of Technology, Newark, NJ 07102, USA
[4] NASA Ames Research Center, Moffett Field, CA 94035, USA
[5] Computer Science Department, New Jersey Institute of Technology, Newark, NJ 07102, USA
[6] Department of Mechanics and Mathematics, Moscow State University, Moscow, 119991, Russia
[7] Moscow Center of Fundamental and Applied Mathematics, Moscow, 119234, Russia
Received 2023 March 22; revised 2023 May 15; accepted 2023 June 1; published 2023 July 27

Abstract

The flux of energetic particles originating from the Sun fluctuates during the solar cycles. It depends on the number and properties of active regions (ARs) present in a single day and associated solar activities, such as solar flares and coronal mass ejections. Observational records of the Space Weather Prediction Center NOAA enable the creation of time-indexed databases containing information about ARs and particle flux enhancements, most widely known as solar energetic particle (SEP) events. In this work, we utilize the data available for solar cycles 21–24 and the initial phase of cycle 25 to perform a statistical analysis of the correlation between SEPs and properties of ARs inferred from the McIntosh and Hale classifications. We find that the complexity of the magnetic field, longitudinal location, area, and penumbra type of the largest sunspot of ARs are most correlated with the production of SEPs. It is found that most SEPs (≈60%, or 108 out of 181 considered events) were generated from an AR classified with the "k" McIntosh subclass as the second component, and these ARs are more likely to produce SEPs if they fall in a Hale class containing a δ component. The resulting database containing information about SEP events and ARs is publicly available and can be used for the development of machine learning models to predict the occurrence of SEPs.

Unified Astronomy Thesaurus concepts: Sunspots (1653); Solar active regions (1974); Solar activity (1475); Solar particle emission (1517); Solar-terrestrial interactions (1473)

1. Introduction

Solar energetic particle (SEP) events are widely known as particle flux enhancements measured by near-Earth satellites. They are among the most dangerous transient phenomena of solar activity because of their negative impacts including health risks for astronauts and airline crews and passengers, damage to satellites and aircraft, radio-wave disturbances, disruptions to power grids, etc. (Kataoka et al. 2018; Martens 2018). Prediction of SEP events before their occurrence could help diminish their impacts by allowing us to take measures ahead of time.

The use of machine learning (ML) to predict SEP events has grown significantly in recent years (Sadykov et al. 2021; Kasapis et al. 2022; Torres et al. 2022), in part due to the increase in the readiness of data related to these events. As an example, in the recent review of 36 SEP prediction models by Whitman et al. (2023), 10 models involved ML. Because the performance of ML may be proportional to the quality and availability of data (Goodfellow et al. 2016), we place significant importance on the development of databases with information about SEP events and active regions (ARs) expanded in time. ARs are sunspots or sunspot groups that represent regions of strong magnetic fields on the solar surface. They are the primary sources of solar flares and coronal mass ejections (CMEs), and consequently of SEPs (Toriumi et al. 2017; Reames 2021). Therefore, it is important to study the link between SEP production and the basic properties of ARs, which include their location on the solar disk, their areas, various magnetic field characteristics, and other measures of AR structure and complexity. Probably the longest available observational measures of AR structure and complexity are the McIntosh (McIntosh 1990) and Hale classifications (Hale et al. 1919). In these classifications, higher classes are assigned to larger and more complex ARs, which are typically more productive in terms of flares, CMEs, and SEPs.

The correlations between SEPs and other solar transient phenomena (flares and CMEs) and McIntosh and Hale classifications were previously studied in several works. For example, Bronarska & Michalek (2017) studied 84 SEP events recorded during the era of the Solar and Heliospheric Observatory spacecraft (1996–2014) with ARs characterized by the McIntosh classification and found that the most energetic SEPs are ejected only from the associated ARs that have a large and asymmetric penumbra. Toriumi et al. (2017) analyzed the Hale classification of ARs that govern large solar flares and eruptions observed between 2010 May and 2016 April. McCloskey et al. (2016) studied the flaring rates and the evolution of ARs in terms of the McIntosh classes using data for solar cycle 22 (SC22). Their results supported the hypothesis that injection of magnetic energy by flux emergence, which results in an increase in the AR class in the

McIntosh system, leads to a higher frequency and magnitude of flare events and thus a greater production of SEP events. The efforts of daily operational probabilistic forecasting of flares and SEPs at the Space Weather Prediction Center (SWPC) at the National Oceanic and Atmospheric Administration (NOAA) also rely on the AR classes (Bain et al. 2021).

The studies mentioned above confirmed the importance of the AR classification and the need for long cross-cycle studies of AR properties and related transient activity. The primary goals of this paper are (i) to develop a homogeneous data set of Hale and McIntosh classes of ARs and associated SEP events spanning a 40 year-long period from 1981 December to 2021 December, which covers the declining phase of SC21, SC22–24, and the rising phase of SC25, and (ii) to perform a direct statistical study of the correlations between SEPs and the Hale and McIntosh classes of ARs using these data. The catalogs utilized in this study are the Solar Region Summary (SRS) records available from the Space Weather Prediction Center (SWPC) NOAA[8] starting from 1996, the United States Air Force (USAF) records of AR classes[9] available from 1981 until 2017, and the catalog of SEP events affecting the Earth maintained by NOAA[10] with the information about proton events available since 1976. The paper is structured as follows. Appendix describes the McIntosh and Hale classifications of ARs. Section 2 describes the catalogs used in this study and related data preparation steps. Section 3 highlights the results of the statistical analysis of the association of AR classes and properties with SEPs, and is followed by the summary of our findings in Section 4. The generated homogeneous data set of AR properties spanning from 1981 until 2021 is publicly available at the Solar Energetic Particle Prediction Portal (SEP^3) webpage.[11]

2. Data Preparation

2.1. Homogeneous Data Set of Solar Active Regions

To maximize the availability of AR classification data for this statistical analysis, we combine the USAF and SWPC NOAA AR catalogs into a continuous AR database covering the 40 yr period from 1981 December to 2021 December. However, the catalogs are not entirely consistent in the way the AR classes are reported. For example, SWPC NOAA SRS records contain information about ARs once per day at 00:00 UT, while the USAF data set can contain several records of the same AR throughout the day from different observing sites (such as the Mount Wilson Observatory and the Boulder Observatory among others) at different times. Moreover, many SEP events originate from regions close to the western limb where information about ARs is either not available or ambiguous. Therefore, we have performed certain steps toward the homogenization of the data sets, namely bringing the records from the USAF data set into a form compatible with the current SWPC NOAA SRS reporting and extrapolating ARs toward the western limb.

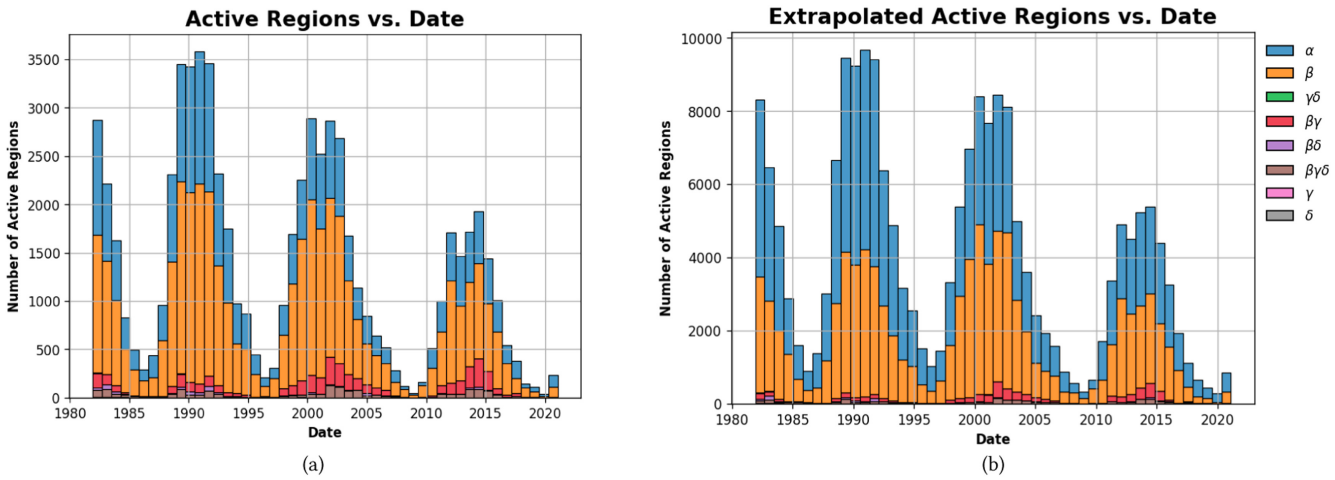
The USAF catalog files were held separately for each year in a text file format. In order to keep consistency with the SWPC NOAA records, we converted their timestamps to follow the format YYYY-MM-DD hh:mm:ss and loaded the records from all files into a single Python Pandas data frame (Pandas Development Team 2020). Then, we sorted the records by AR number and timestamp of an observation, which resulted in groups of ARs that ascended based on their observation times. While inspecting the data set, it was noticed that some entries had an AR area equal to zero, while additional entries for the same AR from other observatories had nonzero areas. We decided that such ARs would take their respective and most recent nonzero values of area, independently of which observatory the information was obtained from.

According to the documentation, the USAF data set followed the Hale classification scheme described in Appendix A.1. Several inconsistencies were found and fixed throughout the catalog. Hale classifications given in Latin letters were changed to Greek letters (*A* was replaced by α, *B* was replaced by β, and so on). Subclasses of β and α (items 1.1, 1.2, 2.1, and 2.2 in Appendix A.1) were designated as β and α, respectively. When a Hale class was considered ambiguous, meaning that its magnetic type was unclear, we inspected the previous or next most recent records of the respective AR and updated the AR record if an appropriate magnetic type for the same AR was found.

The McIntosh classification scheme described in Appendix A.2 is inferred from McIntosh (1990) and the USAF documentation. Inconsistencies in the McIntosh classification were divided into two categories: "ambiguous" and "unambiguous." The McIntosh classes were designated ambiguous when one of the three components was missing, and this missing component had more than one allowed class outlined in Table 2. For instance, the McIntosh classification "F I" was labeled as ambiguous because the penumbra type of the largest spot (second component) was not specified, and based on Table 2, several different classes in the second component are allowed to be paired with "F" as the first component and "I" as the third component. In addition, we have found several McIntosh configurations that are not allowed according to Table 2 and are therefore considered as "ambiguous" (for example, the "HXO" McIntosh class, which was met in the data set). In contrast to the "ambiguous" classifications, an "unambiguous" classification refers to a McIntosh configuration with specifically one or more missing components, which can be unambiguously added to the respective record (for example, the first component "A" allows only "AXX" configuration). For every ambiguous entry, we inspected the most recent previous or next record of the same AR and, if an acceptable McIntosh classification was found, used it for the ambiguous entry.

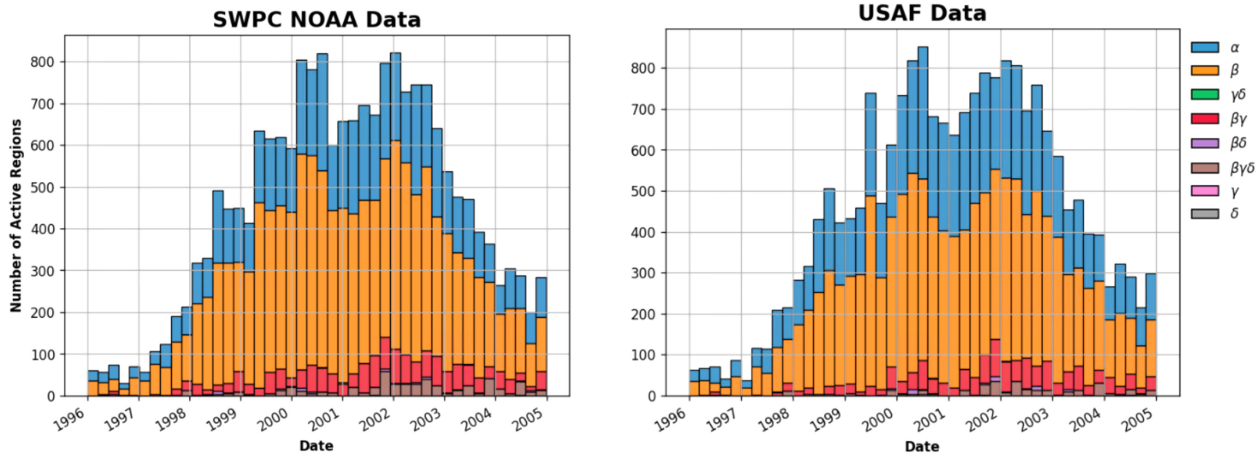
In order to keep the data formatting consistent throughout the merged AR database, each AR record in the USAF catalog had to be approximated to midnight, similar to the SWPC NOAA SRS records. We iterated through each entry in the catalog and approximated its day and time to the next midnight if the entry is the last entry for the day for the given AR number. If there is a subsequent entry the same day with the same AR number, we did not update it but stored its index. Later, saved indices were dropped, keeping only the last and the approximated record of each respective AR during each day. The longitudinal location of the AR record to approximate was updated according to the Carrington rotation rate using the

[8] ftp://ftp.swpc.noaa.gov/pub/warehouse/
[9] https://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features/sunspot-regions/usaf_mwl/
[10] https://umbra.nascom.nasa.gov/SEP/
[11] https://Sun.njit.edu/SEP3/data sets.html



(a) (b)

Figure 1. The continuous AR database was achieved by combining AR catalogs from the Space Weather Prediction Center at the National Oceanic and Atmospheric Administration and the US Air Force Space Weather Wing. The annual AR numbers are visualized through a stacked histogram in panel (a). Panel (b) shows a stacked histogram with the total number of extrapolated ARs vs. date. The legend shows the Hale classifications and the contribution of ARs in a respective Hale class to the total number of ARs in each bin.



(a) (b)

Figure 2. Histograms used to perform the homogeneity test. (a) Data originating from the Space Weather Prediction Center of the National Oceanic and Atmospheric Administration (SWPC NOAA). (b) Data from the US Air Force (USAF). The legend shows the Hale classifications and the contribution of ARs in a respective Hale class to the total number of ARs in each bin. The appearance of each class is consistent throughout the entire figure.

following formula:

$$\text{new longitude} = \text{current longitude} + \left(\frac{\text{difference in hours until midnight}}{24\ \text{hr}}\right) \times 14.2°. \quad (1)$$

By making the modifications described above, we were able to construct the continuous AR data set by combining the data catalogs from the SWPC NOAA and USAF for 1981 December–2021 February. This data set is presented in Figure 1(a). We have also tested the homogeneity of the data provided by the two sources we utilize after applying the processing steps above to Air Force Research Laboratory records. This was done through quantitative comparisons of the number of ARs and their respective features during several overlapping years, from 1997 January 1 to 2004 December 31, during the SC23 maximum. Figure 2 shows the histograms used for our homogeneity test. The histogram on the left was generated using data solely from the SWPC NOAA database, while the histogram on the right was produced using data from the USAF database. Although the total numbers of ARs in each data set are relatively close, the height of several bins corresponding to a specific period of time seems to differ, meaning that the number of ARs recorded by the two solar centers for every single time is not exactly the same. In addition, the test showed that the portions of ARs with $\beta\gamma\delta$ (brown), $\beta\gamma$ (red), and β (orange) magnetic field types also slightly differ in these databases. Thus, we determined that the data sets obtained from the two different solar centers, while generally consistent, are not entirely homogeneous, and this has to be taken into account while utilizing the developed data set for research purposes.

2.2. Linking Records of Solar Energetic Particles and Active Regions

The SWPC NOAA list of SEP events affecting the Earth provides records of the SEP events that, according to the observations of the Geostationary Operational Environmental

Satellite, reached the threshold of 10 particle flux units (pfu) for $\geqslant$10 MeV protons. The current records of SEP events in our possession span six decades, from 1976 April until 2017 September. For the complete list of SEP events please refer to the original source.[12] The list contains information about the start and peak times of each SEP event, the peak flux of $\geqslant$10 MeV protons, and the corresponding information about the record of the preceding CME and soft X-ray flare. The majority of the records also contain information about the location of the host ARs on the solar surface and their NOAA number, allowing us to link this list with the homogeneous AR data set constructed above. We also note here that, although the SEP records provide the date and time of the start and peak, we utilize the time of the preceding flare (more precisely, its peak in 1–8 Å soft X-ray emission) as a reference time for merging the data sets, because the SEP arrival time may vary from minutes to several hours. Inconsistencies were also found in the SEP records and were updated to undertake this statistical analysis. The total number of SEP records in the analyzed list is currently 267. From that number, 36 records did not have the date and time of flare maximum and were removed, leaving 231 SEPs.

Not all SEPs originated from ARs observed on the visible solar disk or within the [−90°, 90°] longitude range. Figure 3(a) shows that several SEP events were identified as originating from ARs behind the western limb with a longitude >90°. The western hemisphere is more directly magnetically connected to the Earth (Parker 1958) and statistical studies of SEP origins demonstrated the asymmetry toward the western limb (Cliver et al. 2020). Therefore, it is important to include information about the ARs located close to the western limb and to map some of the SEP events to these ARs. To accommodate such SEPs, we decided to extrapolate the longitude of all ARs (and their corresponding dates) to cover the entire 360° circumference of the Sun (i.e., until the ARs reach the eastern limb again) by applying the Carrington rotation rate. In order to avoid any duplicates, the AR longitudes were extrapolated only after their last records near the western limb. The extrapolation was performed assuming the Carrington rotation rate according to Equation (1). Figure 1(b) illustrates the continuous AR database after performing the extrapolation. It can be interpreted as a stacked histogram with the total number of extrapolated ARs per year. The legend shows the Hale classifications and the contribution of ARs of a particular Hale class to the total number of ARs in each bin, corresponding to a range of dates. It can be observed from each bin that the number of ARs increased by a factor of approximately three compared to the histogram of ARs in Figure 1(a).

After acquiring the extrapolated AR database and the SEP database, we merged the records into a single data frame containing a one-to-one correspondence between ARs and SEPs generated by ARs. The merging process was performed by considering the corresponding AR numbers and the date of consideration of an AR with the date of the flare peak time recorded for a SEP event (i.e., the class of AR recorded at midnight preceding the time of the flare peak was linked to the flare and, correspondingly, to the SEP records). From the 231 SEP events selected, we were able to match 181 to their AR sources in our extrapolated AR database.

[12] http://umbra.nascom.nasa.gov/SEP/

3. Results and Discussion

3.1. Hale Classification

Figure 3(a) shows the total number of SEP events versus the longitude of the AR producing each one at midnight before initiating the SEP event. The legend shows the contribution of ARs with a particular Hale classification to the number of SEP events within each range of longitudes. Most SEPs are produced in the western hemisphere of the Sun, within the range [0°, 90°]. This characteristic arises because the Earth is magnetically connected to the solar longitudes of ≈75°, meaning that SEPs are more likely to reach Earth if they originate from an AR closer to that range of longitudes. Although this relationship was known and highlighted in previous studies (e.g., Cliver et al. 2020), there are some additional interesting dependences related to the magnetic classes of SEP-producing ARs that we can mention. Table 1 presents the summary statistics of SEP-producing ARs depending on their Hale classes and summarizes the median locations in particular. One can notice from this table that the simpler configurations of Hale ARs (α and β) have their median locations toward the western limb (60° and 30°, correspondingly) with respect to more complex $\beta\gamma$, $\beta\delta$, and $\beta\gamma\delta$ regions (29°, −10°, and 24°). It can also be inferred from Figure 3(a) that a greater number of SEPs came from an α (blue) AR when they were closer to the Earth and the Sun's magnetic connection, while the distribution of SEPs originating from $\beta\gamma\delta$ (brown) ARs shows that a relatively large number of them were generated far from ≈75°. This can be related to the fact that more complex ARs generate stronger flares (in terms of their soft X-ray class) that statistically result in faster and wider CMEs. The CME width was recently indicated to be an important parameter for SEP forecasting (Torres et al. 2022); on the other hand, the recent work by Laitinen et al. (2023) indicated that SEPs can arrive from a wide range of longitudes, even without a wide particle source. We also notice that the AR records close to the western limb and behind it are extrapolated records, and for them the AR class is assumed to be unchanged from the last reliable observations. Correspondingly, we cannot exclude the possibility that SEP records from relatively simple α and β regions close to the western limb can correspond to more complex AR configurations evolved from these α and β regions. It can also be inferred that only six SEP events were found to be generated from behind the western limb.

Figure 3(b) displays the rate of SEP event generation (i.e., a daily climatological probability of the SEP to be produced from an AR of a certain class) in four longitude bins: (−81°, −35°), (−35°, 12°), (12°, 58°), and >58°. The daily climatological probability is calculated as the number of SEPs generated from ARs in a particular Hale class divided by the total number of ARs in our AR database with that Hale class multiplied by 100%, taking into account that each AR has only one record daily. One can notice that the probabilities increase toward western longitudes for both the simpler configurations of ARs (like β) and more complex $\beta\gamma\delta$; both categories of region produce more SEPs when approaching the magnetically connected longitudes.

Figure 3(c) shows a histogram with the number of SEPs versus the area of their respective ARs. It can be generally inferred that ARs with larger areas are more likely to produce SEP events. Table 1 demonstrates that the median areas of

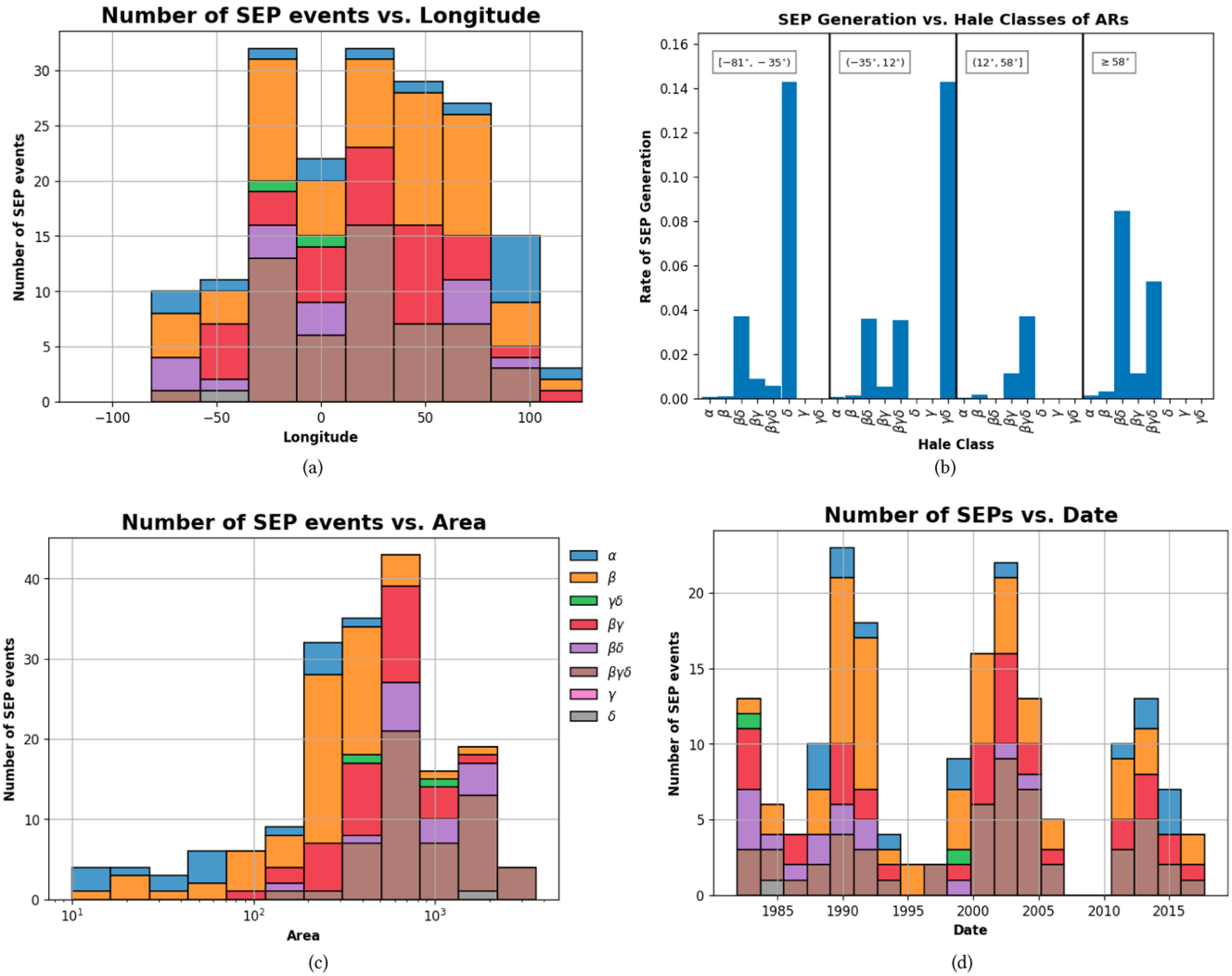


Figure 3. (a) The number of SEPs vs. the location of ARs across the surface of the Sun. The majority of SEPs are seen to be generated in the western hemisphere. The legend shows the Hale classifications and illustrates the contribution of ARs in a particular Hale class to the total number of SEPs in each bin. (b) The rate of production of SEPs by ARs in certain longitude ranges classified by their magnetic field type. (c) The number of SEPs vs. the area of their respective ARs. (d) The number of SEPs within the specific date ranges in each bin. The legend shows the Hale classifications and illustrates the contribution of ARs in a particular Hale class to the total number of SEPs in each bin.

Table 1
Properties of SEP-producing ARs and the Corresponding SEP Events Depending on Their Hale Class

Hale Class	# of SEP Events	Median Longitude of SEP-active ARs[a] (deg)	Median AR Area (10^{-6} th Hemispheres)	Median Peak Flux of SEPs (pfu)
α	16	60.0 ± 37.5	60.0 ± 50.0	91.5 ± 75.5
β	59	30.0 ± 42.0	240.0 ± 130.0	55.0 ± 41.0
$\beta\gamma$	35	29.0 ± 27.0	450.0 ± 210.0	66.0 ± 51.0
γ	0	…	…	…
$\beta\delta$	15	−10.0 ± 49.0	820.0 ± 300.0	180.0 ± 162.0
$\beta\gamma\delta$	53	24.0 ± 32.0	780.0 ± 310.0	134.0 ± 120.0
$\gamma\delta$	2	−12.0 ± 22.0	745.0 ± 405.0	90.0 ± 60.0
δ	1	−43.0 ± 0.0	1870.0 ± 0.0	2500.0 ± 0.0

Notes. Notice that $\gamma\delta$ and δ have poor statistics of SEP events produced from those regions (two events and one event, respectively) and there are no SEP events associated with γ regions in our database.
[a] Determined as the median longitude of the SEP-active ARs at midnight before initiating the SEP event.

SEP-producing ARs vary with their Hale class, and the same is evident in Figure 3(c).

Figure 3(d) shows the number of SEP events versus their respective dates. Comparing with the histograms in Figure 1, we can see that the production of SEP event varies in phase with the 11 yr cycle of solar activity. It also demonstrates the known pattern of solar cycle 24 being weaker than the preceding cycles 23 and 22 in terms of the sunspot numbers,

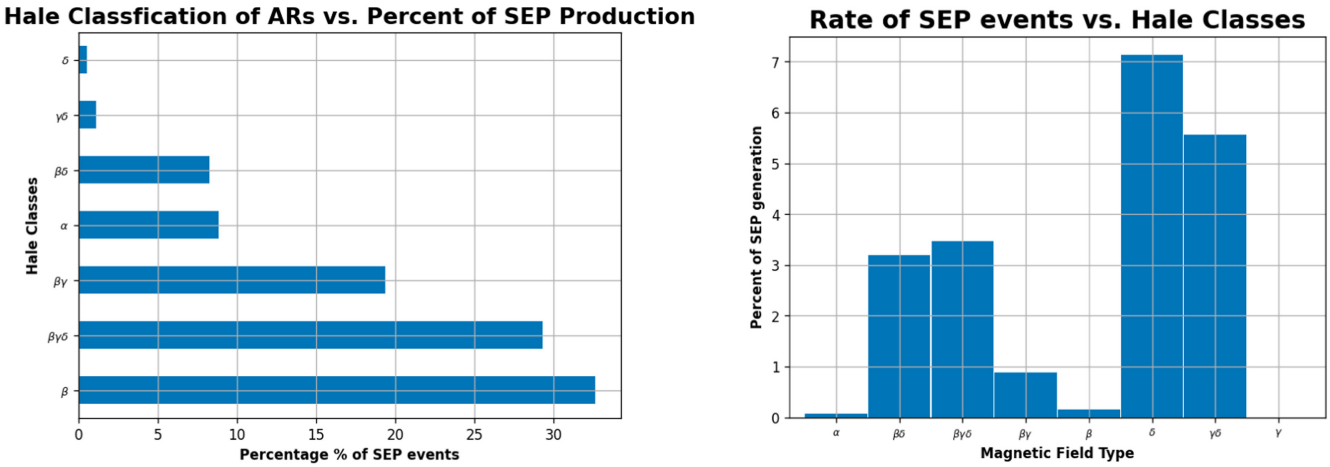


Figure 4. (a) The fraction of the total number of SEP events (as a percentage) that were produced by ARs with a particular Hale class. (b) The rate of production of SEPs by ARs with a particular Hale class. The rate is calculated by dividing the total number of SEPs produced by an AR from a particular Hale class by its corresponding total number of appearances.

Table 2
McIntosh Class Configurations of Sunspot Groups

Class	Penumbra: Largest Spot	Distribution	Number of Configurations (AFWA 2013)
A	x	x	1
B	x	o, i	2
C	r, s, a, h, k	o, i	10
D, E, F	r	o,i	6
D, E, F	s, a, h, k	o, i, c	36
H	r, s, a, h, k	x	5
Total allowed types	⋯	⋯	60

numbers of ARs that appeared at the surface, and the number of SEPs generated.

Figure 4 shows the correlation between the Hale classification of ARs (described in Appendix A.1) and SEP production. Panel (a) shows the percentage of the total number of SEPs produced by ARs classified by a particular Hale class. The percentage was calculated by dividing the total number of SEP events found in the corresponding Hale class by the total number of SEP events in our database, which is 181. Approximately 33% of the total number of SEPs were produced by β ARs, followed by the $\beta\gamma\delta$ class with approximately 29%. The δ ARs appear to have produced only one SEP event, and $\gamma\delta$ have produced only two (see Table 1). Interestingly, even the ARs that are typically assumed not to be active with respect to solar transient events (such as α or β) generated more than 40% of the SEPs considered in this work if combined. This confirms the importance of the characterization of ARs with other parameters in addition to the Hale class, such as McIntosh classes or quantitative magnetic field properties (Sadykov et al. 2021; Kasapis et al. 2022).

ARs exhibiting the δ Hale class in their magnetic field configuration (which are δ, $\beta\delta$, $\gamma\delta$, and $\beta\gamma\delta$) are historically regarded as highly SEP-producing because of their correlation with solar flares. Figure 4(b) shows the rate of SEP production (i.e., the daily climatological probability of SEP events in these regions shown as a percentage on the y-axis) of ARs in a particular Hale class. It can be inferred that the $\beta\gamma\delta$ configurations have a high chance of generating SEP events (approximately 3.5%) given the condition that the $\beta\gamma\delta$ region is observed, confirming what has been concluded historically. The $\beta\delta$ regions show a rate of approximately 3.2%. Although the rates of the $\gamma\delta$ ARs (5.6%) and δ ARs (7.1%) are higher than the rate of $\beta\gamma\delta$ regions, we note that only two and one SEP event records, respectively, exist in our merged data set, and these rates should be interpreted with caution.

3.2. McIntosh Classification

Figure 5 shows the results of a statistical study referencing the McIntosh classification of ARs, described in Appendix A.2, and SEP events. Panel (a) shows the distribution of the total number of SEP events that were produced by ARs with their respective three-component McIntosh classification. Through the legend, we can distinguish the proportion of ARs from a corresponding Hale class that produced SEPs. Thus, it can be inferred that the greatest number of SEP events originated from ARs having "k" as the second McIntosh component. Additionally, the majority of these ARs were classified as $\beta\gamma\delta$, which allows us to close in on a connection between the Hale class ($\beta\gamma\delta$) and the McIntosh subclass (k). Figure 5(b) shows the distribution of SEP events generated from ARs with a particular AR class given by the first component of the McIntosh classification scheme, while Figure 5(c) indicates the total number of SEP events that originated from ARs with a particular distribution of sunspots provided by the first McIntosh component. Because the legend is kept consistent throughout the entire figure, it can be inferred from Figure 5(b) that most SEPs were generated from ARs exhibiting classes E, D, F and corresponding Hale classes $\beta\gamma\delta$, $\beta\delta$, $\beta\gamma$, and β. On the other hand, Figure 5(c) shows almost the same total number of SEPs originating from ARs with distributions of sunspots "i," "c," and "o" with a less even proportion of such ARs in the respective Hale class. The two most obvious correlations are the $\beta\gamma\delta$ regions, which appear to favor the "c" distribution of sunspots, and the β regions favoring the "o" counterpart.

A more defined association of SEP events and McIntosh classes is shown in Figure 6(a). It shows the total number of SEP events generated by ARs with a given penumbra type of the largest sunspot in the AR, defined as the second component

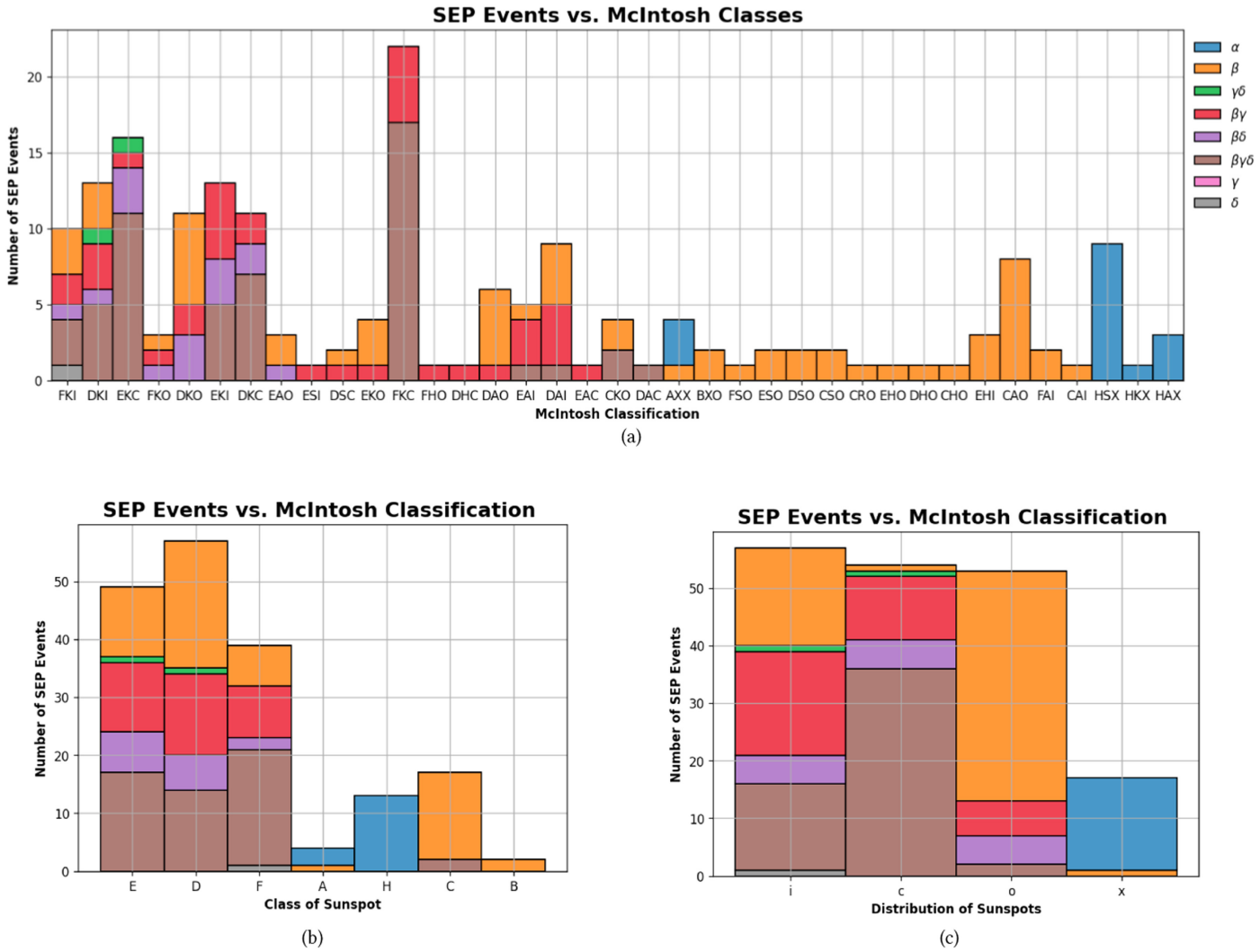


Figure 5. (a) Distribution of SEP events over the entire McIntosh classification scheme. The legend shows the Hale classifications and the contribution of ARs in a particular Hale class to the total number of SEPs in each bin. The appearance of Hale's class is consistent throughout the entire figure. (b) Number of SEPs produced by ARs classified by the first component of the McIntosh scheme, after separating each of the McIntosh components. (c) Number of SEPs produced by ARs classified by the third component of the McIntosh scheme.

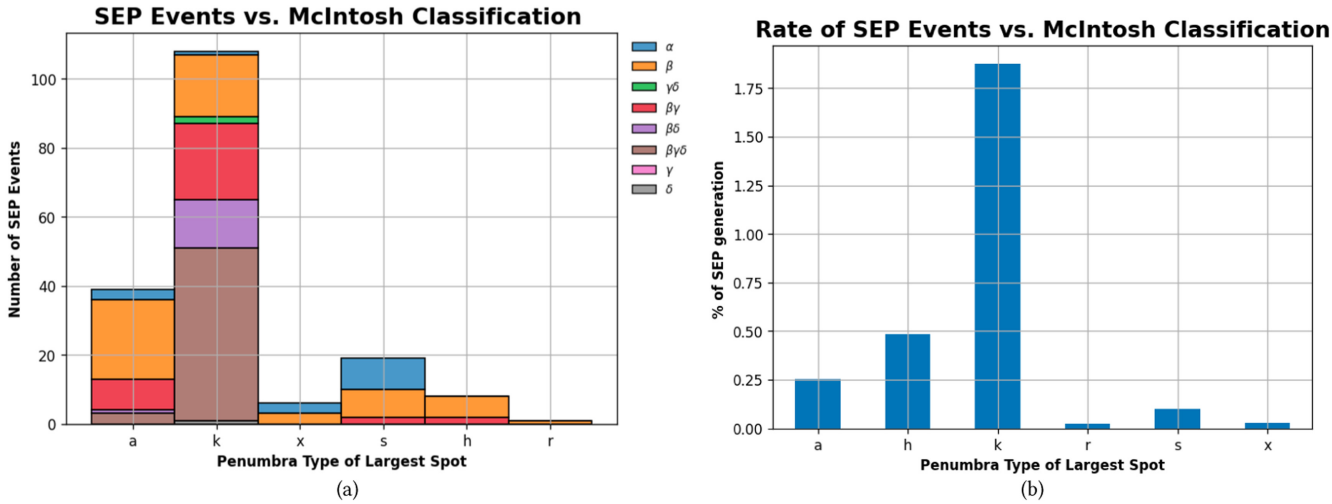


Figure 6. (a) Number of SEP events produced by ARs classified by the second component of the McIntosh classification scheme. The legend shows the Hale classifications and the contribution of ARs in a particular Hale class to the total number of SEPs in each bin. (b) Rate of SEP production of ARs classified by considering the McIntosh component used in panel (a). To calculate the rate, the total number of SEPs generated from ARs with a particular second McIntosh component is divided by the total number of appearances of ARs with that corresponding classification.

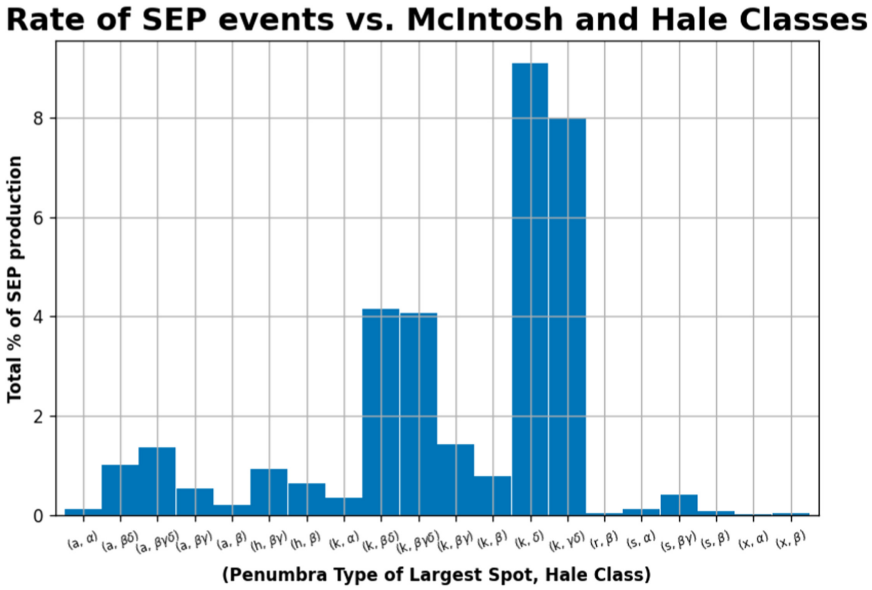


Figure 7. Rate of SEP events (as a percentage) vs. a combination of the second component of the McIntosh classification scheme and the Hale classification scheme. When combined, the rates of SEP production are found to be greater for either or both classes in consideration than when considered separately. Two major groups of ARs are found to produce the most SEPs: (k, δ) and (k, $\gamma\delta$) ARs.

of a McIntosh class. Based on this figure, we can conclude that ARs with a "k" subclass as the second component of a McIntosh classification produce the most SEPs. In addition, it can be inferred that ARs with $\beta\gamma$, $\beta\delta$, $\gamma\delta$, and $\beta\gamma\delta$ magnetic field configurations favor a "k" subclass as the second component of a McIntosh class in terms of SEP production. Figure 6(b) shows the rate of production of SEPs from ARs with a particular penumbra type of their corresponding largest sunspot. The rate was calculated as in Section 3.1 for Hale classes. The number of SEPs generated from ARs in a particular class is divided by the total number of ARs in our AR database with that class multiplied by 100%. Following the results, it can be concluded that ARs with the "k" subclass as the second McIntosh component have a higher rate of SEP production, and consequently they are more likely to produce SEPs than ARs with any other class in this component.

3.3. Combined McIntosh and Hale Classifications

Figure 6 shows that some ARs are more likely to produce SEPs when their second McIntosh component is determined as "k" and paired with certain Hale class configurations. To further study this relationship, we combined the two classification schemes, and ARs were grouped based on two features: the Hale class and the second component of the McIntosh class. The results are shown in Figure 7. We can conclude that the rate of SEP production is highest for ARs with any given Hale class combined with the "k" subclass as the second component of a McIntosh classification. In addition, it is important to note that these rates are higher than the rates found by separately considering Hale classes and the considered subclass of the McIntosh classification. For example, the rates of SEP events produced by $\beta\gamma\delta$ ARs and ARs with "k" second McIntosh components were $\sim$3.5% and $\sim$1.8% in that order and as shown in Figures 4(b) and 6(b). Because (k, $\beta\gamma\delta$) ARs have a $\sim$4.1% rate of SEP production, they are more likely to produce SEPs than ARs that fall into only one of the "k" or $\beta\gamma\delta$ class categories. The same can be concluded for β, $\beta\delta$, $\beta\gamma$, $\gamma\delta$, and δ ARs if they are also determined to have the "k" McIntosh subclass. These results indicate the importance of including both Hale and McIntosh classes for advancing the forecasting of SEPs.

4. Summary

We presented a statistical study of the correlation between SEP events and properties, and Hale and McIntosh classes of ARs. Properties of ARs included their longitude and area, while classes came from the Hale and McIntosh classification schemes. We concluded that longitudes closer to the magnetic connection between the Sun and the Earth are more favorable for less complex ARs, such as α and β, in terms of SEP production. ARs of larger areas, irrespective of their Hale class, are also more likely to produce SEP events. The total number of SEPs produced by ARs of a certain Hale or McIntosh class is not a sufficient characteristic to determine the hazard a given AR may represent. For instance, although $\beta\delta$ ARs generated only 15 SEP events throughout the entire time frame in consideration, they have a higher rate of producing SEP events (about 3.2%) than β regions that produced the most SEP events (59). It was found that most SEPs were generated from an AR classified with the "k" McIntosh subclass as the second component ($\approx$60%, or 108 out of 181 considered events), and some of these ARs are more likely to produce SEPs if they fall in a certain Hale class. For example, $\beta\gamma\delta$ ARs have a $\sim$3.5% chance of producing SEPs, while ARs with a "k" as the second McIntosh component have a $\sim$1.8% rate of SEP production. An AR with a combination of these two classes has a probability of $\sim$4.1% of producing a SEP event. The developed homogeneous data set of AR classes spanning more than three solar cycles (1981–2021) can be utilized for the development of robust forecasting approaches for SEP events and other transient solar activity validated over an extensive time period and varying solar activity.

Acknowledgments

This research was supported by NASA Early Stage Innovation program grant No. 80NSSC20K0302, NASA LWS grant No. 80NSSC19K0068, NSF EarthCube grant No.

1639683, and NSF grant No. 1835958. V.M.S. acknowledges the NSF FDSS grant No. 1936361 and NSF grant No. 1835958. E.I. acknowledges the RSF grant No. 20-72-00106.

Appendix
Description of Active Region Classification

A.1. Hale Classification of Active Regions

The Hale et al. (1919) classification of ARs describes the configuration of their magnetic fields (see also Jaeggli & Norton 2016). In this classification, unipolar sunspot groups belong to α class (or type), the β type includes distinct bipolar ARs, complex magnetic field configurations with an irregular distribution of polarities constitute the γ type, and ARs of δ type contain sunspots that within one penumbra contain opposite magnetic polarities (umbrae) separated by less than two heliographic degrees. Combinations of these primary classes can be used to describe complex ARs. For example, a complex bipolar configuration with several distinct opposite polarities is described as $\beta\gamma$ type. If such an AR contains a *delta*-type sunspot it is classified as $\beta\gamma\delta$ (He et al. 2021). Below, we summarize the classes and subclasses presented in the USAF data set (AFWA 2013; Roslan 2018).

1. ALPHA (α). A unipolar sunspot or sunspot group, surrounded by a plage of opposite magnetic polarity without visible sunspots.
   (a) ALPHA p (α). A unipolar sunspot or sunspot group surrounded by a plage of the same polarity and *followed* by an elongated plage or faculae of the opposite polarity (this subclass is used in the USAF classification only).
   (b) ALPHA f (α). A unipolar sunspot or sunspot group surrounded by a plage of the same polarity and *preceded* by an elongated plage or faculae of the opposite polarity (USAF only).
2. BETA (β). A distinct bipolar group with magnetic field balance between the separated leading and trailing sunspots.
   (a) BETA p (β). A bipolar sunspot group, in which the magnetic field strengths and areas of the *leading* sunspot are dominant (USAF only).
   (b) BETA f (β). A bipolar sunspot group, in which the magnetic field strengths and areas of the *trailing* sunspot are dominant (USAF only).
3. BETA–GAMMA ($\beta\gamma$). A complex sunspot group of several opposite-polarity sunspots without a well-defined dividing line between regions of opposite polarity.
4. GAMMA (γ). A sunspot group with irregular (intermixed) magnetic structures of opposite polarities.
5. BETA–DELTA ($\beta\delta$). A β-type AR, in which a sunspot umbra contains opposite polarities within the same penumbra.
6. BETA–GAMMA–DELTA ($\beta\gamma\delta$). A $\beta\gamma$-type AR, in which a sunspot umbra contains opposite polarities within the same penumbra.
7. GAMMA–DELTA ($\gamma\delta$). A γ-type AR, in which a sunspot umbra contains opposite polarities within the same penumbra.
8. DELTA (δ). A sunspot group with a sunspot umbra containing opposite polarities (separated by less than two heliographic degrees) within the same penumbra.

A.2. McIntosh Classification of Active Regions

McIntosh (1990) modified and expanded a classification scheme based on white-light images of sunspot groups, which was initially developed by Cortie (1901). It is composed of 17 parameters describing the structure of sunspot groups, combinations of which provide 60 allowed classes (McCloskey et al. 2016). The McIntosh classification is represented by three letters *Zpc*, where *Z* is the modified Zurich class, *p* is the type of the largest sunspot, and *c* is the degree of compactness of the sunspot group. A more detailed description of the classification scheme based on the USAF database documentation (AFWA 2013) is provided below.

A.2.1. Modified Zurich Class—Z

The modified Zurich class is based on the appearance of the sunspot penumbra and the extent of the group (McIntosh 1990). A judgment of the complexity of the sunspot group is not required for this component of the McIntosh scheme. It includes seven classes:

A. Unipolar group of sunspots without penumbrae with the size of three heliographic degrees or less.
B. Bipolar group of sunspots without penumbrae with the size of three heliographic degrees or greater.
C. Bipolar sunspot group, in which only spots of one polarity have a penumbra.
D. Bipolar sunspot group, in which sunspots of both polarities have a penumbra, and the group size is less than 10 heliographic degrees.
E. Bipolar sunspot group, in which sunspots of both polarities have a penumbra, and the group size is greater than 10 but does not exceed 15 heliographic degrees.
F. Bipolar sunspot group, in which sunspots of both polarities have a penumbra, and the group size is greater than 15 heliographic degrees.
G. Unipolar group of sunspots with a penumbra, in which the principal (usually leading) sunspot is remaining from an old bipolar group.

A.2.2. Penumbra of Largest Spot—p

Six classes in this component described the appearance and the structure of the largest sunspot penumbra:

a. No penumbra.
b. Rudimentary or incomplete irregular penumbra that is usually brighter than a mature penumbra and has a granular or mottled (not filamentary) fine structure.
c. Small mature circular or elliptical penumbra (or a compact cluster of umbrae near the sunspot center) with a fine filamentary structure with a diameter of 2.5 heliographic degrees or less.
d. Small mature asymmetric (not circular or elliptical) penumbra with filamentary fine structure with a diameter of 2.5 heliographic degrees or less, usually containing two or more umbrae within it.
e. Large mature symmetric penumbra with a diameter greater than 2.5 heliographic degrees (with the corresponding AR area greater than ~250 millionths of the solar hemisphere).
f. Large mature asymmetric penumbra with a diameter greater than 2.5 heliographic degrees (with the

corresponding AR area greater than ∼250 millionths of the solar hemisphere).

A.2.3. Sunspot Distribution—c

The third component describes the density of sunspots in a sunspot group. It provides additional information about the AR size and potentially indicates the presence of a strong magnetic field gradient across the polarity inversion line between opposite-polarity sunspots.

a. Undefined for single sunspots and unipolar sunspot groups.
b. "Open" sunspot group with few, if any, "interior" spots (usually small spots without penumbra or pores) between the leading and trailing sunspots.
c. "Intermediate" sunspot group with many interior spots between the leading and trailing sunspots, but none of the interior spots has a mature penumbra.
d. "Compact" sunspot group populated with many strong interior sunspots with at least one mature penumbra; in extreme cases, the entire group can be embedded in one continuous penumbral area.

A summary of the McIntosh classification is given in Table 2 (AFWA 2013). The total number of possible classes is 60. However, not every combination of the components listed above is permitted.

ORCID iDs

Russell D. Marroquin https://orcid.org/0000-0002-3364-7463
Viacheslav Sadykov https://orcid.org/0000-0002-4001-1295
Alexander Kosovichev https://orcid.org/0000-0003-0364-4883
Irina N. Kitiashvili https://orcid.org/0000-0003-4144-2270
Gelu M. Nita https://orcid.org/0000-0003-2846-2453
Egor Illarionov https://orcid.org/0000-0002-2858-9625
Aatiya Ali https://orcid.org/0000-0003-3196-3822

References

AFWA 2013, Air Force Weather Agency Manual, Air Force Weather Agency, https://ngdc.noaa.gov/stp/space-weather/online-publications/miscellaneous/afrl_publications/afwaman15-1_space-environmental-observations.pdf
Bain, H. M., Steenburgh, R. A., Onsager, T. G., & Stitely, E. M. 2021, SpWea, 19, e2020SW002670
Bronarska, K., & Michalek, G. 2017, AdSpR, 59, 384
Cliver, E. W., Mekhaldi, F., & Muscheler, R. 2020, ApJL, 900, L11
Cortie, A. L. 1901, ApJ, 13, 260
Goodfellow, I., Bengio, Y., & Courville, A. 2016, Deep Learning (Cambridge, MA: MIT Press)
Hale, G. E., Ellerman, F., Nicholson, S. B., & Joy, A. H. 1919, ApJ, 49, 153
He, Y., Yang, Y., Bai, X., et al. 2021, AdAst, 2021, 5529383
Jaeggli, S. A., & Norton, A. A. 2016, ApJL, 820, L11
Kasapis, S., Zhao, L., Chen, Y., et al. 2022, SpWea, 20, e2021SW002842
Kataoka, R., Sato, T., Miyake, S., Shiota, D., & Kubo, Y. 2018, SpWea, 16, 917
Laitinen, T., Dalla, S., Waterfall, C. O. G., & Hutchinson, A. 2023, A&A, 673, L8
Martens, P. C. 2018, LPI Contribution, 2063, 3188
McCloskey, A. E., Gallagher, P. T., & Bloomfield, D. S. 2016, SoPh, 291, 1711
McIntosh, P. S. 1990, SoPh, 125, 251
The pandas development team 2023, pandas-dev/pandas: Pandas v2.0.2, Zenodo, doi:10.5281/zenodo.3509134
Parker, E. N. 1958, ApJ, 128, 664
Reames, D. V. 2021, Solar Energetic Particles. A Modern Primer on Understanding Sources, Acceleration and Propagation (Lecture Notes in Physics) (Cham: Springer)
Roslan, M. A. Z. B. 2018, MSc thesis, Science Univ. of Malaysia
Sadykov, V., Kosovichev, A., Kitiashvili, I., et al. 2021, arXiv:2107.03911
Toriumi, S., Schrijver, C. J., Harra, L. K., Hudson, H., & Nagashima, K. 2017, ApJ, 834, 56
Torres, J., Zhao, L., Chan, P. K., & Zhang, M. 2022, SpWea, 20, e2021SW002797
Whitman, K., Egeland, R., Richardson, I. G., et al. 2023, AdSpR, in press