

Statistical Analyses of Solar Active Region in SDO/HMI Magnetograms detected by Unsupervised Machine Learning Method DSARD

R. CHEN,^{1,2} W. LU,¹ Q. HAO,^{1,3} Y. MENG,¹ P. F. CHEN,^{1,3} AND C. SHI¹

¹*School of Astronomy and Space Science, Nanjing University, Nanjing 210023, China*

²*School of Mathematics, Nanjing University, Nanjing 210023, China*

³*Key Laboratory of Modern Astronomy and Astrophysics (Nanjing University), Ministry of Education, Nanjing 210023, China*

ABSTRACT

Solar active regions (ARs) are the places hosting the majority of solar eruptions. Studying the evolution and morphological features of ARs is not only of great significance to the understanding of the physical mechanisms of solar eruptions, but also beneficial for the hazardous space weather forecast. An automated DBSCAN-based Solar Active Regions Detection (DSARD) method for solar ARs observed in magnetograms is developed in this work, which is based on an unsupervised machine learning algorithm called Density-Based Spatial Clustering of Applications with Noise (DBSCAN). The method is then employed to identify ARs on the magnetograms observed by the Helioseismic and Magnetic Imager (HMI) onboard Solar Dynamics Observatory (SDO) during solar cycle 24 and the rising phase of solar cycle 25. The distributions of the number, area, magnetic flux, and the tilt angle of bipolar ARs in latitudes and time intervals during solar cycle 24, as well as the butterfly diagram and drift velocities are obtained. Most of these statistical results based on our method are in agreement with previous studies, which also guarantees the validity of the method. In particular, the dipole tilt angles in ARs in solar cycle 24 and the rising phase of solar cycle 25 are analyzed which reveal that 13% and 16% of ARs, respectively, violate Hale’s law.

Keywords: Solar active regions — Solar magnetic fields — Solar cycle — Astronomical techniques

1. INTRODUCTION

Active regions (ARs) are the patchy volumes in the solar atmosphere with strong magnetic fields, where a variety of activities occur, manifesting in complex spatial and temporal behaviors extending from the photosphere to the corona (van Driel-Gesztelyi & Green 2015). Sunspots, being the initial features discernible as tracers of ARs on the Sun, appear dark areas on the surface of the Sun in continuum, occasionally perceptible to the naked eyes (Solanki 2003). They mark the locations where magnetic field is sufficiently strong to suppress the convective flows that transport heat from the interior to the solar surface. Observations of sunspots in ARs have revealed several characteristics that are crucial to understanding the physical mechanisms of the solar cycles. One such example is the equatorward migration of

sunspots known as the “Butterfly Diagrams” (Maunder 1904), which further led to the discovery of various magnetic properties of sunspots, including the famous Joy’s law and Hale’s polarity law (Maunder 1904). The most significant aspect of ARs and sunspots are their capacity to generate strong solar flares and, sometimes, coronal mass ejections (CMEs), which have the potential to significantly impact the space weather. Consequently, the study of the evolution of ARs and their relationship with solar eruptions is not only useful in predicting eruptive events (Schrijver 2007; Chen 2011; Bobra & Couvidat 2015; Barnes et al. 2016; Sun et al. 2022; Zhang et al. 2022), but also in understanding the long-term evolution of solar activities (Hathaway 2010; Ravindra et al. 2022).

The National Oceanic and Atmospheric Administration (NOAA) Space Weather Prediction Center (SWPC) has developed a widely referenced catalog of active regions (ARs) based on white-light images. NOAA assigns a unique number to an AR if it contains at least one sunspot, with the numbering determined either manu-

Corresponding author: Q. Hao
haoqi@nju.edu.cn

³ R. Chen and W. Lu are co-first authors.

ally or through computer-aided inspection assisted by human operators. This catalog offers detailed information, including the location of sunspot groups, their areas, and magnetic morphology.

However, significant advancements in observational techniques have led to a substantial increase in both the quantity and quality of multiwavelength observational data. Traditional manual detection or labeling of ARs on images has become increasingly inadequate in meeting the demands for rapid and accurate processing of large datasets, particularly for long-term statistical studies and solar dynamo modeling. Recognizing the limitations of human inspection, the development of automated detection methods has gained momentum. These methods aim to provide objective and consistent extraction of ARs, addressing the challenges posed by manual processes and enabling more efficient analysis of solar data.

An early class of automated detection methods for ARs was primarily developed using image processing techniques (Zharkov et al. 2005; McAteer et al. 2005a,b; Benkhalil et al. 2006; Colak & Qahwaji 2008; Zhang et al. 2010; Higgins et al. 2011; Cui et al. 2021). Verbeeck et al. (2013) conducted a comprehensive comparison of four typical automated detection methods for ARs or sunspots observed in images with different wavelengths. These methods typically involve threshold segmentation on magnetograms or white-light images, followed by region-growing and morphological operations to derive the areas of ARs. For instance, Zharkov et al. (2005) employed an edge detection method after standardizing images to identify sunspot candidates. This was followed by local thresholding for image segmentation, and subsequent watershed and morphological operations to smooth features and fill gaps. The temporal variations in the detected sunspot areas showed a high correlation with NOAA’s records, validating the effectiveness of the proposed method. This sunspot detection approach was applied to white-light images captured by the Michelson Doppler Imager (MDI; Scherrer et al. 1995) aboard the Solar and Heliospheric Observatory (SOHO) spacecraft from 1996 to 2004. The results revealed that the number of sunspots increases near-exponentially as the area decreases over the observed period. Additionally, the sunspot areas exhibited a periodicity of approximately 7–8 years in terms of north-south asymmetry.

Zhang et al. (2010) developed an automated AR detection method incorporating three image processing techniques: intensity thresholding, region-growing, and morphological operations. This method was tested and applied to the Carrington synoptic magnetograms con-

structed from SOHO/MDI images spanning 1996 to 2008. The method identified 1730 ARs and yielded significant statistical insights, including morphological features, magnetic flux, and drift velocities during solar cycle 23. The true positive and false positive rates relative to the NOAA AR catalog were found to depend on the detection threshold. For example, the frequency distribution of ARs in terms of area size and magnetic flux follows a log-normal function, but shifts to a power-law distribution when the detection thresholds are lowered. Despite computational biases, the authors suggested that the distribution of solar magnetic features, from large to small, might consist of two components: a log-normal distribution for large ARs and a power-law distribution for smaller ARs and other minor features. However, these methods rely on empirically assigned thresholds and struggle to effectively separate ARs in complex magnetic field regions on high-resolution images, leading to incomplete identification. Their robustness depends on thresholds determined through extensive trial and error, making them susceptible to instability due to detector sensitivity degradation.

The advancement of machine learning methods in computer vision technology has significantly expanded their application in the detection of solar ARs. One widely used approach is based on clustering, a form of unsupervised machine learning. Barra et al. (2008) introduced a fuzzy clustering technique to automatically segment EUV images into regions of coronal holes, quiet Sun, and active regions. Noticing some artifacts of the results, they later improved the method and applied it to the entire set of EIT solar images over solar cycle 23 (Barra et al. 2009), yielding results consistent with earlier studies. Caballero & Aranda (2013) applied clustering methods to group segmented regions into active regions for EUV observations via the region growing technique. A comparative study of three clustering techniques revealed that hierarchical clustering delivered the best performance.

Jiang et al. (2022) introduced a novel approach by first applying the scale-invariant feature transform (SIFT) to extract different regions. These regions were then clustered based on their density peaks, enabling the automatic detection of active regions in magnetograms. A key advantage of this method over the region-growing approach is that it eliminates the need for seed point selection, thereby avoiding the issue of multiple neighboring active regions being incorrectly identified as a single AR.

It is worth noting that Turmon et al. (2002) pioneered a Bayesian image segmentation technique, leveraging statistical models trained on scientist-provided labels to

characterize the quiet Sun, faculae, and sunspots. This technique was further extended and refined in a subsequent study (Turmon et al. 2010). The method was applied to identify active region patches observed by the Helioseismic and Magnetic Imager (HMI; Schou et al. 2012) onboard the Solar Dynamics Observatory (SDO; Lemen et al. 2012). This led to the creation of the widely used Spaceweather HMI Active Region Patch (SHARP; Bobra et al. 2014), which provides detailed maps containing information on magnetic fields, Doppler velocity, and continuum intensity.

Convolutional neural networks (CNNs) represent another widely utilized machine learning method, standing at the forefront of feature segmentation in the machine learning domain. Supervised learning algorithms typically train models by adjusting parameters based on labeled data until the desired performance is achieved. Quan et al. (2021) employed two object detection models, Fast R-CNN and YOLO-V3, to detect the bounding boxes of ARs in magnetograms observed by SDO/HMI. The experimental results demonstrated that both methods exhibit high detection accuracy and rapid processing speeds. While bounding boxes around ARs may be effective for flare prediction, focusing on the detailed characteristics of ARs reveals that bounding boxes alone are insufficient. To accurately extract specific AR regions within these boxes, intensity threshold processing or pixel-scale classification, as proposed by Turmon et al. (2010), is necessary. Semantic segmentation models, such as U-Net (Ronneberger et al. 2015), have proven effective in accurately identifying regions of interest for automatically detecting filaments/prominences and ARs with high precision (Zhu et al. 2019; Liu et al. 2020, 2021; Zhang et al. 2024). However, unlike other solar activities, ARs in magnetograms often appear as diffuse, fragmented areas of dots and patches. This indicates that while semantic segmentation models can delineate fragmented patches of ARs, an additional step is required to determine which fragments belong to a single AR. One potential solution is to employ an instance segmentation model, as previously implemented for filaments by Guo et al. (2022). However, the number of AR fragments significantly exceeds that of filaments, complicating the data labeling process.

Despite the higher accuracy and generalization ability of these supervised learning models, they require large amounts of objective and consistent labeled data of ARs, which is difficult to free from the subjectivity of human inspection. In contrast, unsupervised learning algorithms do not require large amounts of data for feature extraction but instead on an appropriate algorithm according to the characteristics of the data it-

self. In this paper, we employ Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, a classical density-based spatial clustering unsupervised learning algorithm (Schubert et al. 2017), to develop an automated detection method for solar active regions observed in magnetograms. The method was then employed to automated detect the ARs observed by SDO/HMI during solar cycle 24 and the rising phase of solar cycle 25, and a statistical analyses of AR features was conducted. The unsupervised automated detection method is described in Section 2, followed by a description of its performance. The statistical results of ARs detected by aforementioned method are presented in Section 3 and compared with the published results in Section 4, where we also draw several conclusions.

2. METHOD

2.1. Data Acquisition

The present study employs the line-of-sight (LOS) full-disk magnetograms observed by SDO/HMI. The SDO/HMI dataset has been acquiring LOS magnetograms at a 45 s cadence, with a spatial resolution 1 arcsec (with each image possessing 4096×4096 pixels) since the SDO launch on 2010 February 11. The dataset under consideration is in the form of a single image selected per day, spanning the period from May 2010 to December 2023, and is acquired through the Joint Science Operations Center¹.

2.2. DSARD: DBSCAN-based Solar Active Regions Detection

An automated detection method for solar active regions observed in magnetograms was developed, and this method was named DSARD (DBSCAN-based Solar Active Regions Detection). The method is based on the Density-Based Spatial Clustering of Applications with Noise (DBSCAN; Schubert et al. 2017), a popular clustering algorithm used in data analysis and machine learning to identify arbitrarily shaped clusters in imaging data. A significant advantage of DBSCAN over other clustering methods is its ability to effectively handle noise and outliers by categorizing points based on their density, eliminating the need to predefine the number of clusters. In essence, DBSCAN can classify magnetic field regions into distinct active regions without requiring prior knowledge of their number.

The core principle of DBSCAN relies on two key parameters: the distance criterion ϵ and the minimum number of pixel points required for a region to be con-

¹ <http://jsoc.stanford.edu/HMI/Magnetograms.html>

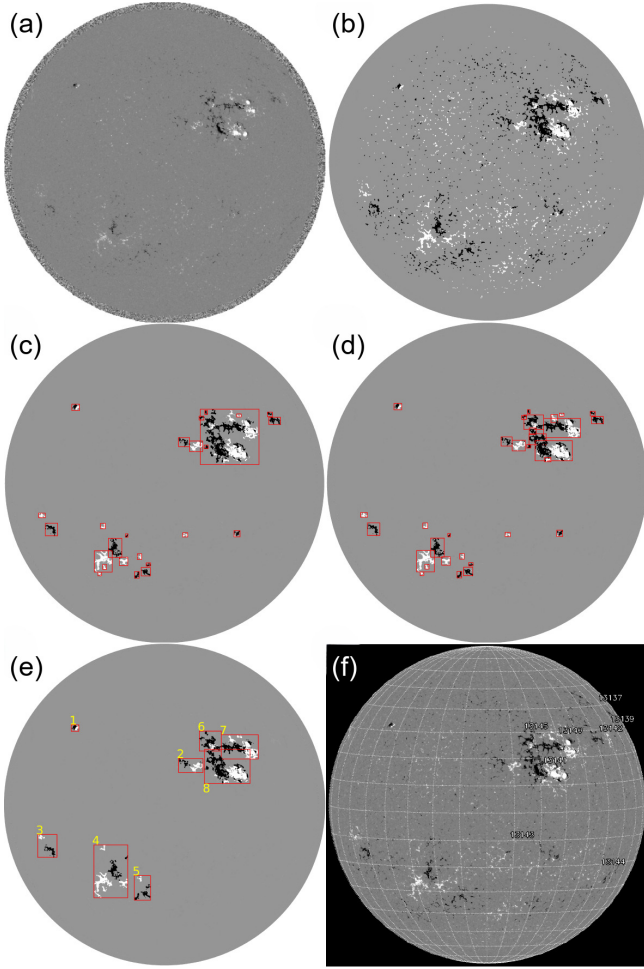


Figure 1. An observation taken on November 12, 2022 as an example to show the processing modules of the automated detection system. (a) The original observation obtained by SDO/HMI. (b) The magnetogram after threshold segmentation. (c) The detected result after the first (global) DBSCAN clustering, with each red box representing a cluster. (d) The detected result after the second (reformative) DBSCAN clustering. (e) The final detected AR results after filtering, merging, and labeling the clusters obtained from DBSCAN². Each red box with a number represents a bounding rectangle of the identified solar AR. (f) NOAA-identified solar active regions in the same day obtained from the Solar-Monitor website (<https://solarmonitor.org>) for comparison.

considered dense, denoted by P_{min} . Based on these parameters, pixel points are categorised into three types:

- **Core Point:** A point with at least P_{min} points within its ϵ -neighborhood.
- **Border Point:** A point that has fewer than P_{min} points within its ϵ -neighborhood but is in the neighborhood of a core point.
- **Noise Point:** A point that is neither a core point nor a border point.

After categorizing the points, clustering begins by identifying all core points and grouping them into clusters based on their reachability through other core points. Border points are then assigned to the nearest cluster, while noise points remain unclassified. This approach excels at discovering clusters of varying shapes and sizes and is robust against noises, making it particularly suitable for astronomical applications where traditional methods may fail.

The DSARD algorithm conceptualizes solar active regions in magnetograms as concentrations of pixels with relatively higher magnetic field intensities. This premise guided the initial application of grayscale threshold segmentation, consistent with previous methods (Turmon et al. 2002; Zhang et al. 2010; Turmon et al. 2010). However, DSARD goes beyond simple thresholding by considering both pixel connectivity and density distribution, leading to the implementation of DBSCAN. It is acknowledged that the performance of DBSCAN heavily depends on the selection of its hyperparameters. To address this, a double iteration of DBSCAN is employed to filter out noise points and generate clusters of varying shapes, as detailed below. The DSARD method consists of the following four steps:

1. **Threshold-based Image Segmentation:** The analysis commences with the implementation of a thresholding technique on the image, thereby accentuating pixels exhibiting substantial magnetic field intensities. The *threshold* is set to be 150 Gauss, a level that encompasses a substantial proportion of pertinent pixels, analogous to the threshold employed for SHARP data (Bobra et al. 2014). The outcome of this segmentation process is depicted in Figure 1(b), wherein positive and negative magnetic fields are delineated by white and black patches, respectively. However, the analysis reveals a pervasive distribution of noise across the solar disk.
2. **First global DBSCAN:** In the first DBSCAN, a smaller value is assigned to ϵ and P_{min_1} in order to filter out noise points remaining after threshold segmentation, thus resulting in the formation of preliminary clusters. It is evident that some clusters are larger than the true clusters due to the initial broad parameter settings, as illustrated in Figure 1(c). The relatively larger cluster on the northwest side, which includes several active regions, is clearly visible. A second DBSCAN is employed to address this issue.
3. **Second reformative DBSCAN:** In the subsequent DBSCAN processing, clusters that exceed a predefined size threshold, designated as *maxSize*, are designated as excessively large. In practice, *maxSize*

was set to 3000 Mm^2 . Consequently, the second DB-SCAN was implemented with larger values for P_{min_2} . This step facilitates the segmentation the oversized clusters into more precise size. As Demonstrated in Figure 1(d), the substantial cluster observed on the northwestern region of the solar disk in Figure 1(c) has been meticulously divided into multiple smaller clusters.

4. Integration: Following the procedure described above, relatively scattered clusters are obtained. As illustrated in Figure 1(d), some clusters are found to be relatively small and/or to comprise unipolar regions. These regions should be integrated to form a complete active region. Initially, a parameter *ratio* is set and the clusters are divided into three categories based on it. The definition of a “positive cluster” is such that the number of points with positive magnetic polarity in the cluster exceeds the number of points with negative polarity by the factor of *ratio*. The definition of a “negative cluster” is such that the number of points with negative magnetic polarity in the cluster exceeds the number of points with positive polarity by the factor of *ratio*. Clusters that do not fit these two categories are referred to as “neutral clusters”.

Next, the distance between the closest points of any two clusters on the solar sphere is calculated. Due to computational constraints, the nearest points between two clusters on the image are identified, and their distance on the solar surface is approximated to be the distance between the two clusters. This distance is then multiplied by a coefficient α , determined by the cluster categories (positive, negative, or neutral), to define a measurement of the distance between clusters. Notably, the distance coefficient between a positive and a negative cluster is relatively small. The values of α are summarized in Table 1.

For each cluster, the nearest cluster is identified. If their distance is less than the minimum distance *minDistance* (set to 50 Mm), they are merged. This process repeats until no further merging is possible. Additionally, neutral clusters with a pixel count below *minSize* = 75 Mm^2 and any “positive” or “negative” clusters with a pixel count below $4 \times \text{minSize}$ = 300 Mm^2 are discarded. The factor of four reflects the observational fact that these clusters often represent decayed active regions, where one polarity region is more diffuse or attenuated. After this process, clusters of appropriate size with reasonably matched positive and negative polarities are identified as active regions, as illustrated in Figure 1(e).

Table 1. Coefficient α based on cluster attributes.

α	Positive	Negative	Neutral
Positive	0.5	0.5	1
Negative	0.5	0.5	1
Neutral	1	1	2

Table 2. Parameters in our method.

Parameter	value
<i>threshold</i>	150 Gauss
ϵ	30 pixels
P_{min_1}	200 pixels
P_{min_2}	500 pixels
<i>maxSize</i>	3000 Mm^2
<i>ratio</i>	10
<i>minDistance</i>	50 Mm
<i>minSize</i>	70 Mm^2

The advanced stages of segmentation and integration not only enhance the robustness of the algorithm to parameter variations but also effectively incorporate the morphological features of solar active regions into the computational framework, leveraging the magnetic intensity information from SDO/HMI magnetograms. The detailed parameters are summarized in Table 2. For comparison, an example image with active regions identified by NOAA on the same day is shown in Figure 1(f). In addition to NOAA ARs 13140, 13141, and 13145, our method identified five additional ARs, labeled as No. 1 to 5. NOAA ARs 13142, 13143, and 13144 were excluded due to their weak and diffuse nature. Besides, although our method accounts for the projection effects, it still fails if the active regions are near the solar limb, such as NOAA ARs 13139, 13142, and 13144.

2.3. Performance

The effectiveness of the above-mentioned identification algorithm has been rigorously validated, showcasing notable advantages in two key areas. First, the primary strength of the DSARD method lies in its exceptional sensitivity to the data density, enabling the detection of smaller or more diffuse ARs that might be overlooked by other algorithms or even manual identification methods. This capability is demonstrated in Figures 1(e) and 1(f), where the DSARD method identified ARs labeled No. 1 and 3 to 5 in the northeast and southeast quarters of the solar disk—regions not marked by NOAA. Second, the sophisticated integration of connectivity and density considerations in our algorithm allow for a more refined segmentation of adjacent ARs. This approach differs significantly from classical image processing tech-

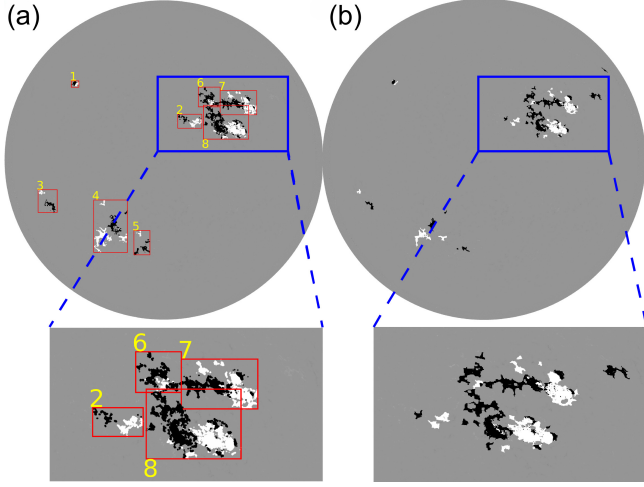


Figure 2. An example to compare the detected AR results by the present method and the classical image processing method. (a) The results detected by the present method are shown in the top panel, which are equivalent to those illustrated in Figure 1(e). The lower panel displays an enlarged section of the blue box region. (b) The results detected by the classical image processing method.

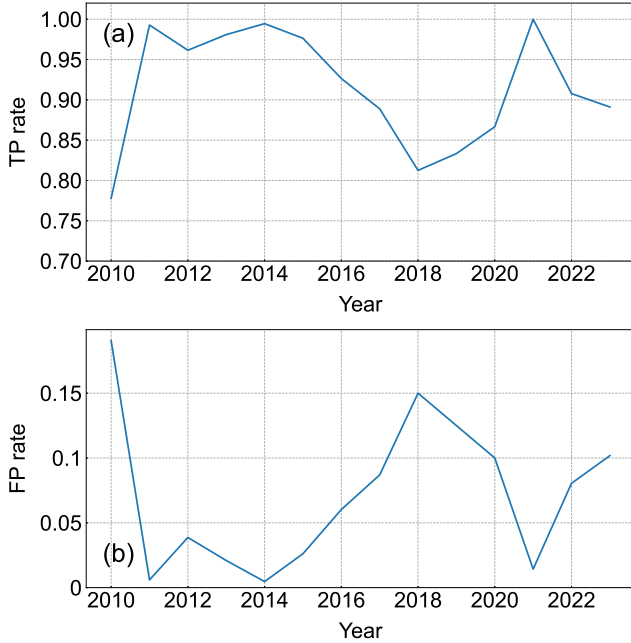


Figure 3. Variation of true positive rate R_{TP} (a) and false positive rate R_{FP} (b) from 2010 to 2023.

niques, such as region-growing methods, which tend to indiscriminately group all connected areas. Figures 2(a) and (b) provide a comparison between the identification results of the DSARD method and those of the region-growing method. The region-growing method fails to capture some nuances of ARs and incorrectly merges

distinct ARs due to its inherent limitations. While morphological open operations are often used to disconnect linked ARs, they result in smaller detected AR areas, as shown in Figure 2(b). In contrast, the DSARD method effectively distinguishes complex ARs while preserving their boundaries, as illustrated in the enlarged panel of Figure 2(a).

Although the algorithm calculates distances at the pixel scale, resulting in slightly longer processing times compared to conventional methods, the processing speed remains acceptable, i.e., approximately 30 seconds per 4096×4096 pixel image on a standard personal computer. The DSARD method and the detected AR features are available on GitHub². Additionally, copies of the method and data are archived on Zenodo³.

To quantitatively evaluate the performance of our method, the NOAA catalog⁴ serves as the “ground truth”. We processed 4984 full-disk magnetograms from May 2010 to December 2023, detecting 25,178 ARs. The detection process is applied to a full-disk magnetogram per day, meaning the same active region (AR) may be detected multiple times. To address the issue of duplicated detections and minimize the projection effects, we focus on a longitudinal range of $\pm 6^\circ$ from the central meridian of the solar disk. Within the specified range, the final number of ARs is 2863. Detailed AR counts can be found in Table 3. For comparison, we only consider the NOAA-labeled ARs within the specified longitudinal range. The true positive rate R_{TP} and the false positive rate R_{FP} are then defined, following the approach of Zhang et al. (2010):

$$R_{TP} = \frac{N_{TP}}{N_{TP} + N_{FN}} = \frac{N_{TP}}{N_{NOAA}}, \quad (1)$$

$$R_{FP} = \frac{N_{FP}}{N_{FP} + N_{TN}} = \frac{N_{FP}}{N_{FP} + N_{NOAA}}, \quad (2)$$

where N_{TP} represents the number of true positives, defined as cases where the center locations reported by NOAA fall within the minimum bounding rectangle of the detected ARs; N_{FP} denotes the number of false positives, where detected ARs are not recorded in the NOAA catalog; N_{FN} refers to the number of false negatives, indicating ARs in the NOAA catalog that the automated method fails to detect; N_{TN} , the number of true negatives, cannot be determined since there is no concept of a “negative AR event”. As shown in Figure 3, the

² <https://github.com/Chenruishuo/DSARD>

³ <https://doi.org/10.5281/zenodo.14222292>

⁴ <http://solarcyclescience.com/activeregions.html>.

true positive rate R_{TP} and false positive rate R_{FP} vary from 2010 to 2023, as indicated by the panels (a) and (b), respectively. The average R_{TP} and R_{FP} are 91.8% and 7.2%, respectively. It is notable that the relatively low R_{TP} is evident in the years 2010 and around 2018, both of which are due to the scarcity of ARs near solar minimum.

The yearly variation of the numbers of ARs detected by the proposed method and from the NOAA catalog during 2010 – 2023 is illustrated in Figure 4. While the distribution of AR numbers detected by the proposed method follows a similar trend to that of the NOAA catalog, the former consistently identifies more ARs. The discrepancies in AR counts are more pronounced in the southern hemisphere compared to the northern hemisphere, particularly around year 2014. This increased detection of ARs can be attributed to the enhanced sensitivity of the DBSCAN-based method to the data density, enabling it to identify smaller or more diffuse ARs that may be overlooked by other methods.

3. RESULTS

As outlined in Section 2.3, the method employed in this study processes the SDO/HMI line-of-sight magnetograms on a daily basis from May 2010 to December 2023, covering most of solar cycle 24 and the rising phase of solar cycle 25. To avoid multiple detections and minimize projection effects, a longitudinal range limit of $\pm 6^\circ$ from the central meridian of the solar disk is applied. Within this range, 1991 ARs were detected during solar cycle 24 and 872 ARs during the rising phase of solar cycle 25. The detailed counts of the detected ARs are summarized in Table 3. In addition to generating butterfly diagrams and analyzing drift velocities, the geometric center (location), area, magnetic flux, and tilt angles of bipolar ARs are calculated for different years and latitude bands. The statistical results are presented in detail in the following subsections.

3.1. Distributions of the Number of Active Regions

As mentioned in Section 2.3, Figure 4 reveals a strong correlation between the yearly variation of AR numbers detected by our method and those listed in the NOAA catalog from 2010 to 2023. Both datasets indicate the presence of two peaks in AR numbers during solar cycle 24, with the first occurring around 2012 and the second, more pronounced peak around 2014. Notably, the first peak is dominated by ARs in the northern hemisphere, while the second peak is dominated by those in the southern hemisphere, as illustrated in Figures 4(b) and (c). During the solar maximum of cycle 24, the monthly AR count exceeded 20, but it declined to less

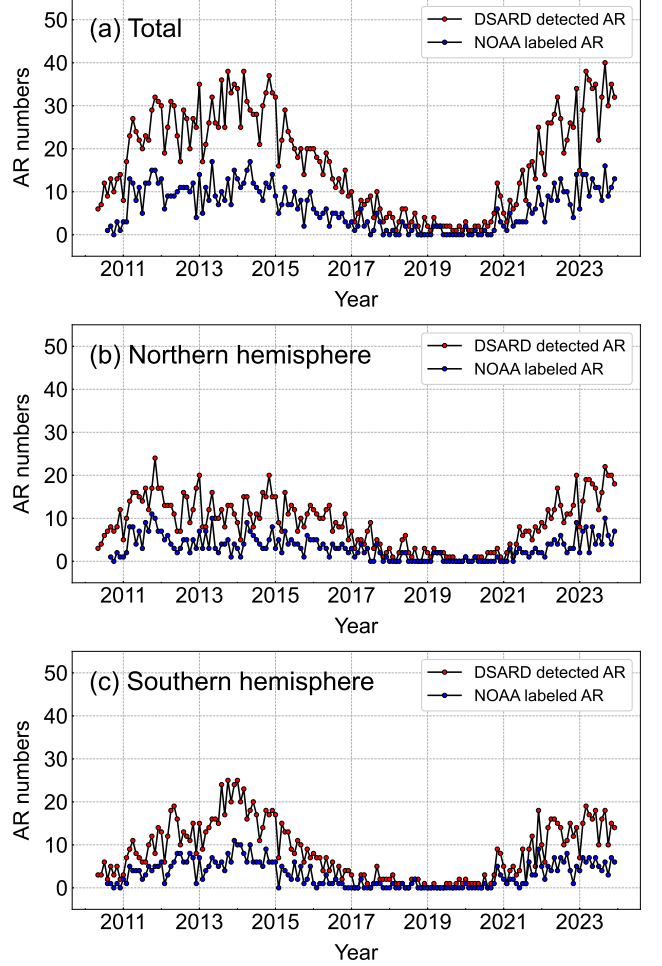


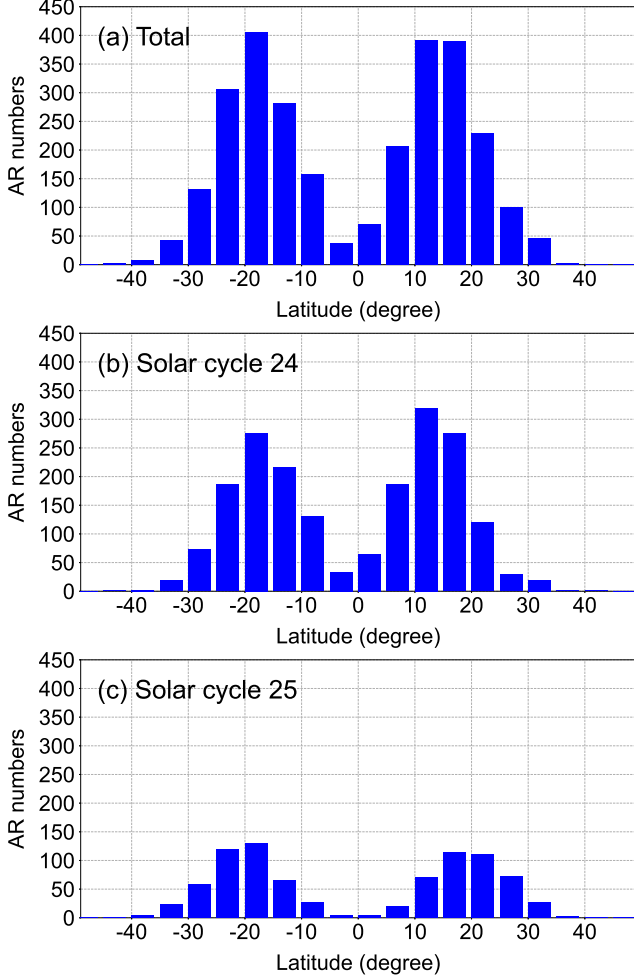
Figure 4. Distribution of the AR number detected by our method and NOAA catalog from 2010 to 2023. (a) Distribution of the total AR number from 2010 to 2023. Each dot represents the sum of the AR number per month. Panels (b) and (c) are similar to panel (a) but for the number of ARs in the northern and southern hemispheres, respectively.

than 5 or even 0 per month during the declining phase from 2018 to 2020. In contrast, during the rising phase of solar cycle 25, the AR numbers increased rapidly and are significantly higher compared to those of solar cycle 24, suggesting that the strength of the rising phase of solar cycle 25 may surpass that of solar cycle 24.

The latitudinal distribution of ARs is illustrated in Figure 5, showing a bimodal distribution with ARs primarily concentrated in the latitude band of 10° – 20° . Few ARs are observed at latitudes higher than 40° . The analysis indicates that during solar cycle 24, the peak in the southern hemisphere occurs in the latitude band 15° – 20° , while in the northern hemisphere, it is in the 10° – 15° band. In contrast, during the rising phase of solar cycle 25, the peaks appear in the 15° – 20° band in both hemispheres. Notably, during solar cycle 24, there

Table 3. Automatically detected AR numbers from 2010 to 2023.

Time interval	Solar Cycle	Data numbers	Detected			AR	Numbers	
			Total	North	South		$\pm 6^\circ$	
2010 May – 2019 Dec	24	3,523	17,523	9,267	8,256	1,991	1,037	954
2020 Jan – 2023 Dec	25	1,461	7,655	3,798	3,875	872	430	442

**Figure 5.** Histograms of active regions showing their distribution with latitude. (a) Total events; (b) Active regions in solar cycle 24; (c) Active regions in the rising phase of solar cycle 25.

are more active regions in the northern hemisphere than in the southern hemisphere, and it is reversed during the rising phase of solar cycle 25.

3.2. Butterfly Diagram and Drift Velocity variations of Active Regions

The latitudinal distribution of ARs over a solar cycle forms a “butterfly diagram”, a pattern known as “Spörer’s Law of Zones” (Hathaway 2010). This latitudinal migration toward the equator reflects the meridional flow associated with the solar dynamo (Choudhuri

2021). The butterfly diagram for ARs during solar cycle 24 and the rising phase of solar cycle 25 is presented in Figure 6. Most ARs are located below 40° , with only a few exceptions. The relatively large ARs are predominantly found in the low-latitude band of 10° – 15° in both hemispheres. During solar cycle 24, the ARs in the northern hemisphere appear for a longer time compared to those in the southern hemisphere. Specifically, ARs in the northern hemisphere emerged from 2010 and persisted until the end of solar cycle 24 in 2020, gradually decreasing in both area and number each year. In contrast, ARs in the southern hemisphere increased in number and area from 2010, peaked around 2014, and then declined rapidly by 2016, with only a few appearing after 2018. Relatively large ARs began to appear around 2012 in the northern hemisphere, while in the southern hemisphere, they emerged since late 2012 and became particularly large and numerous in 2014 and 2015. This suggests that ARs in the southern hemisphere were stronger than those in the northern hemisphere during these years. During the rising phase of solar cycle 25, ARs began to emerge more intensely than in solar cycle 24. A notable feature of this phase is the earlier appearance and larger size of ARs in the southern hemisphere compared to the northern hemisphere. Additionally, there was a gradual increase in both the number and area of ARs from late 2021 to 2023.

To conduct a quantitative analysis of AR migration during solar cycle 24 and the rising phase of solar cycle 25, the average latitude of ARs is calculated at 30-day intervals. These average latitudes are then fitted using a cubic polynomial. Figure 6(b) illustrates the evolution of the monthly average latitudes of ARs during solar cycles 24 and 25, with blue and orange dots representing the respective cycles. The blue and orange lines correspond to the cubic polynomial fits of the monthly average latitudes, while the dashed vertical line marks the start of solar cycle 25. Due to the dispersed latitude distribution around 2020, data from this period are excluded from the analysis. The monthly average latitudes of ARs in both hemispheres begin at approximately 20° , which are lower than the latitudes shown in the butterfly diagram of all ARs. This discrepancy arises because the majority of ARs are located below 25° , as demonstrated

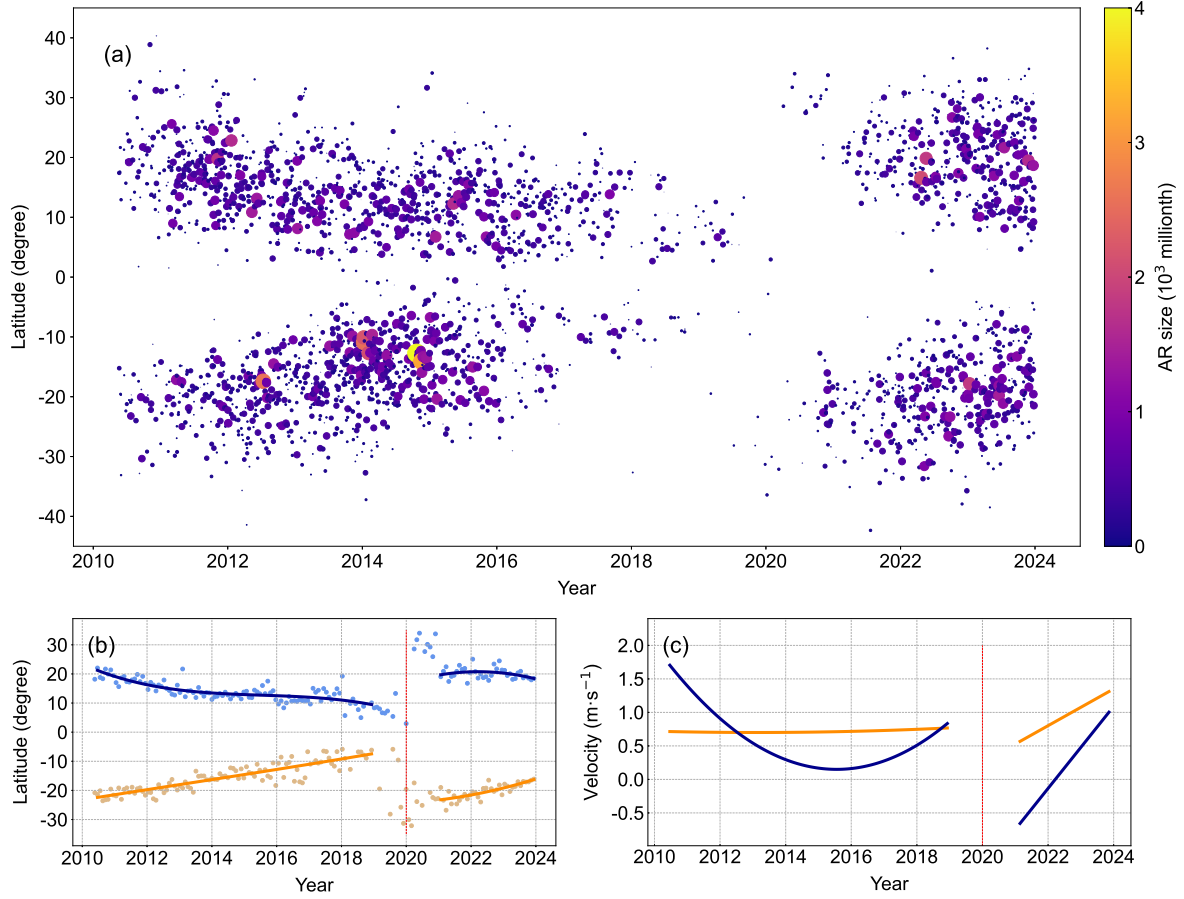


Figure 6. Butterfly diagram and latitudinal migration of ARs during the solar cycle 24 and the rising phase of solar cycle 25. (a) Butterfly diagram of active regions from 2010 to 2023. Each dot represents a single AR, with the size and color of the dot corresponding to the respective area of the active region. (b) Temporal evolution of the monthly average latitude of active regions in the northern (blue dots) and southern (orange dots) hemispheres is illustrated, with the blue and orange lines representing the cubic polynomial fitted lines of the monthly average latitude of active regions. The dash vertical line indicates the beginning of solar cycle 25. (c) Fitted drift velocity variations in the northern (blue line) and southern (orange line) hemispheres, respectively.

by the histograms of the latitudinal distribution of AR numbers in Figure 5.

As shown in Figure 6(c), the derived drift velocity evolutions in the northern and southern hemispheres are represented by blue and orange curves, respectively. During solar cycle 24, ARs in both hemispheres migrated toward the equator, displaying significant asymmetry. ARs in the southern hemisphere exhibited relatively stable drift velocity variations. A decline in drift velocities was observed during the rising phase around 2014, followed by a modest increase toward the end of solar cycle 24, with maximum and minimum velocities of 0.8 m s^{-1} and 0.7 m s^{-1} , respectively. In contrast, the drift velocities of ARs in the northern hemisphere showed greater variability compared to those in the southern hemisphere. A decline from 1.5 m s^{-1} to 0.1 m s^{-1} was observed in late 2015, followed by an increase to 0.8 m s^{-1} by the end of solar cycle 24 in 2019.

The rising phase of solar cycle 25 also manifested clear asymmetries in drift velocities between the two hemispheres. In the northern hemisphere, a deceleration of velocities toward relatively high latitudes occurred from 2021 to 2022, followed by an acceleration toward the equator, reaching 1.0 m s^{-1} in 2023. In contrast, the southern hemisphere demonstrated a consistent acceleration toward the equator, with velocities increasing from 0.5 m s^{-1} to 1.2 m s^{-1} .

3.3. Distributions of Active Region Areas

The area of ARs is measured in millionths, defined as millionths of the total solar surface area. AR areas are categorized into four groups based on their size: <50 millionths, 50–250 millionths, 250–500 millionths, and >500 millionths. The proportions of these groups are approximately 1 : 2 : 4 : 2. Figure 7(a) illustrates the yearly distributions of AR numbers of these four area

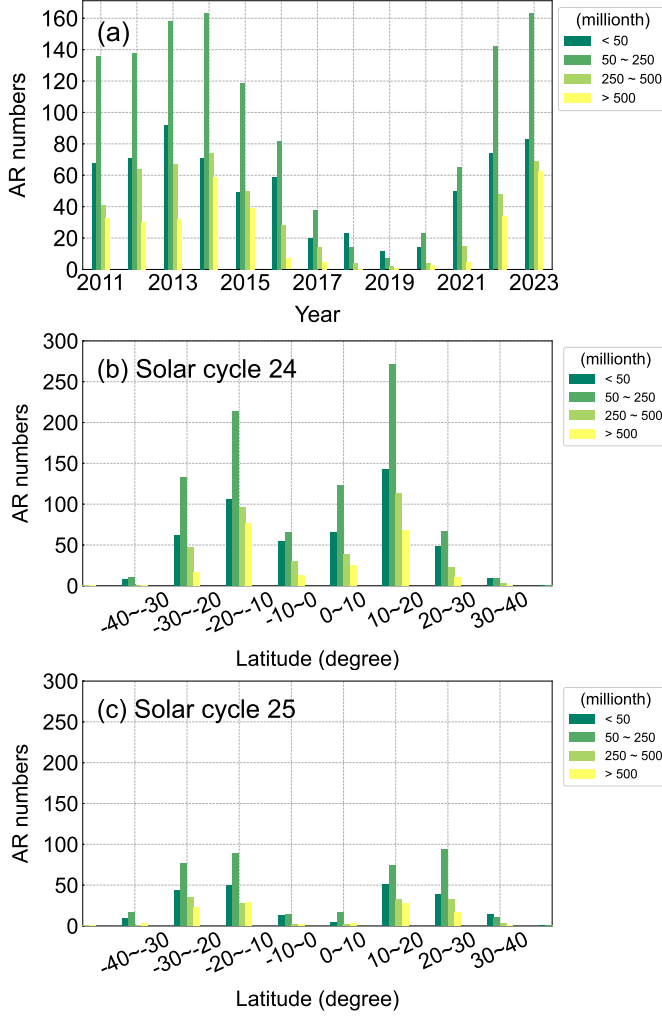


Figure 7. Distribution of AR numbers in different area ranges with respect to latitude and year from 2010 to 2023. (a) Distribution of AR numbers in different area ranges with respect to year. (b) Distribution of AR numbers in different area ranges with respect to latitude during solar cycle 24. Panel (c) is similar to (b) but for that during the rising phase of solar cycle 25.

groups. In general, the number of ARs increased from the beginning of the solar cycle to the solar maximum and then decreased toward the end of solar cycle 24. The peak in AR numbers occurred in 2014, except for the group with areas < 50 millionths, which shows two peaks in 2013 and 2016 during solar cycle 24. During the rising phase of solar cycle 25, the number of ARs in all four groups increased steadily from 2020 to 2023.

As illustrated in Figures 7(b) and 7(c), the latitudinal distribution of AR numbers of the four area groups during solar cycle 24 and the rising phase of solar cycle 25 exhibits a bimodal pattern. During solar cycle 24, the peaks are observed in the latitude band of 10° – 20° in both hemispheres. Within the latitude band of 0° –

20° , the AR numbers in the four area groups are higher in the northern hemisphere compared to the southern hemisphere. Conversely, in the latitude band of 20° – 30° , the southern hemisphere shows higher AR numbers. In contrast, during the rising phase of solar cycle 25, no significant hemispheric asymmetry is observed. The peaks in AR numbers predominantly occur in the 10° – 20° latitude band, with the exception of ARs in the northern hemisphere with areas of 50–250 millionths and those in the southern hemisphere with areas of 250–500 millionths. Notably, the AR numbers in the 30° – 40° latitude band during the rising phase of solar cycle 25 are comparatively higher than those in solar cycle 24, suggesting a broader distribution of ARs at relatively higher latitudes in solar cycle 25.

As shown in Figures 8(a), 8(c), and 8(e), the yearly and latitudinal distribution of cumulative active region (AR) areas reveals a prominent peak in the late part of 2014, consistent with the AR number trends shown in Figure 7(a). The cumulative AR areas exhibit a more rapid increase during the rising phase of solar cycle 25 compared to solar cycle 24. Additionally, the cumulative AR areas across different latitude bands display a bimodal distribution, as illustrated in Figures 8(c) and 8(e). The largest cumulative areas are observed in the latitude band of 10° – 20° in both hemispheres. Notably, the latitudes of the peaks during the rising phase of solar cycle 25 are higher than those observed in solar cycle 24.

3.4. Distributions of Active Region Magnetic Flux

The magnetic flux is calculated by summing the magnetic fluxes threading through the pixels within each AR. This process yields the positive, negative, and total unsigned fluxes for each individual AR. Similar to the classification of AR areas, the magnetic fluxes of ARs are categorized into intensity-based groups: $< 10^{21}$ Mx, $1\text{--}5 \times 10^{21}$ Mx, $5\text{--}10 \times 10^{21}$ Mx, and $> 10^{22}$ Mx. The yearly distributions of the positive, negative, and total unsigned fluxes for these four magnetic flux groups are shown in Figures 9(a) and (b). Most ARs with positive and negative fluxes are below 5×10^{21} Mx, while the total unsigned fluxes are predominantly below 10^{22} Mx. The peak flux for ARs with fluxes less than 10^{21} Mx occurs in 2013, whereas ARs with larger fluxes ($1\text{--}5 \times 10^{21}$ Mx) peak one year later, in 2014, during solar cycle 24. For ARs with fluxes within $5\text{--}10 \times 10^{21}$ Mx, two peaks are observed in 2012 and 2014, though the first peak is less pronounced. ARs with magnetic fluxes exceeding 10^{22} Mx appear exclusively during the rising phase and extend into the years around the solar maximum of solar cycle 24. Notably, ARs of all four magnetic flux groups

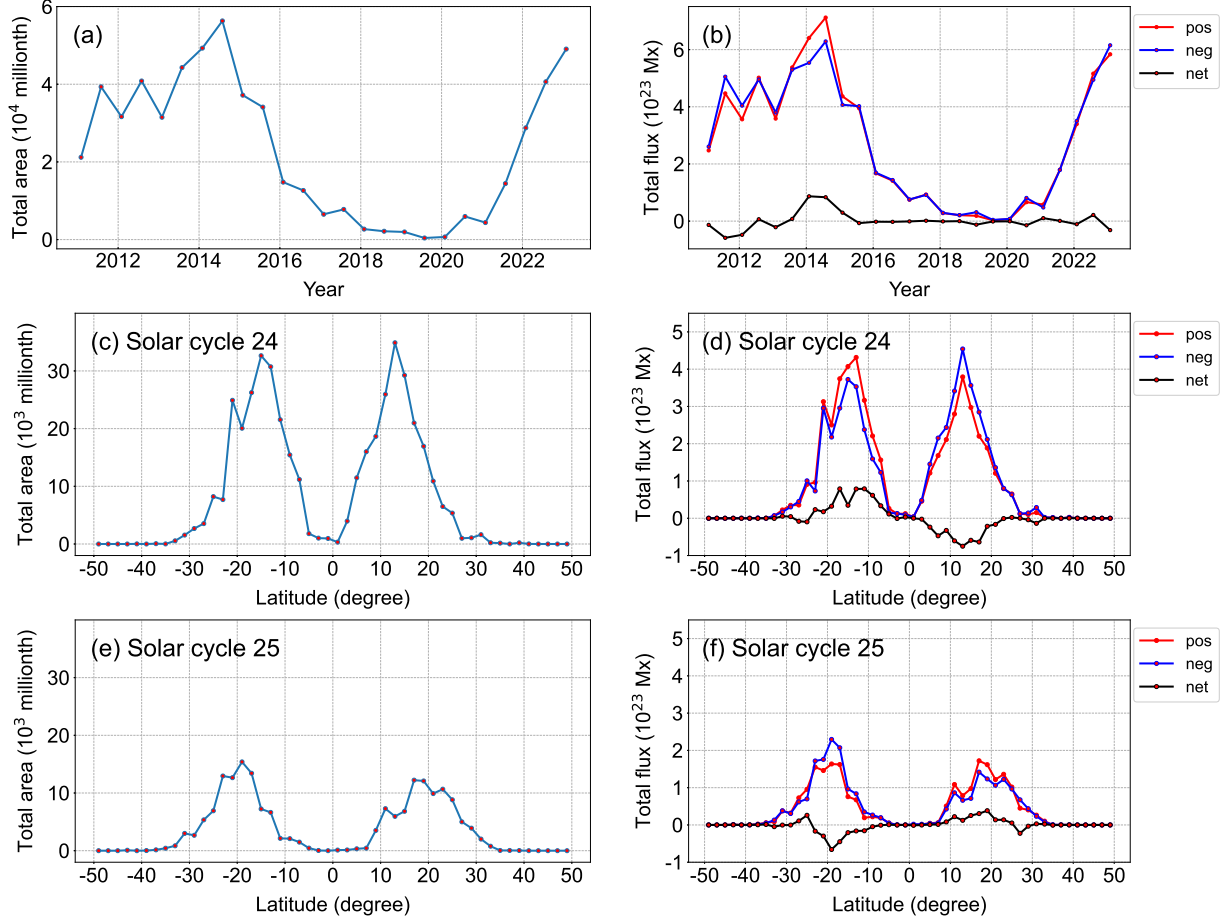


Figure 8. Distribution of cumulative area and magnetic flux with respect to latitudes and years.

exhibit an increasing trend during the rising phase of solar cycle 24.

Figures 9(c)–(f) illustrate the latitudinal distribution of ARs with positive, negative, and total unsigned fluxes across four magnetic flux groups during solar cycle 24 and the rising phase of solar cycle 25. The distribution of ARs in these flux groups is consistent with the distribution of AR areas. The latitudinal distribution of AR magnetic flux exhibits a bimodal pattern, with peaks in the latitude band of 10° – 20° in both hemispheres. ARs with larger magnetic fluxes (exceeding 10^{22} Mx) are predominantly found at lower latitudes, below 30° . During solar cycle 24, ARs with magnetic fluxes below 10×10^{21} Mx are more numerous in the northern hemisphere compared to the southern hemisphere. However, no significant hemispheric asymmetry is observed during the rising phase of solar cycle 25. Additionally, ARs with relatively small fluxes (less than 1×10^{21} Mx) show a predominance of positive flux in the southern hemisphere and negative flux in the northern hemisphere.

Figures 8(b), (d), and (f) display the yearly and latitudinal distribution of the cumulative magnetic flux of

ARs. The largest cumulative values of positive, negative, and net magnetic fluxes are observed in 2014, consistent with the trends in AR numbers and AR areas during solar cycle 24. The similar variation trends for cumulative positive and negative magnetic fluxes, as well as cumulative areas, arise from their strong correlation. Notably, the cumulative areas increased more rapidly during the rising phase of solar cycle 25 compared to solar cycle 24. The cumulative positive and negative magnetic fluxes across different latitude bands also exhibit a bimodal distribution, as shown in Figures 8(c) and (b), with peaks occurring in the latitude band of 10° – 20° in both hemispheres. The net magnetic flux is negative from 2010 to 2013, becomes positive with relatively large values in 2014, and remains near zero during the decay phase of solar cycle 24. During the rising phase of solar cycle 25, the positive and negative magnetic fluxes show a rapid increase, while the net magnetic flux remains approximately zero. The latitudinal distribution of the cumulative net magnetic flux of ARs reveals that, during solar cycle 24, the net magnetic flux is positive in the southern hemisphere and negative in the northern

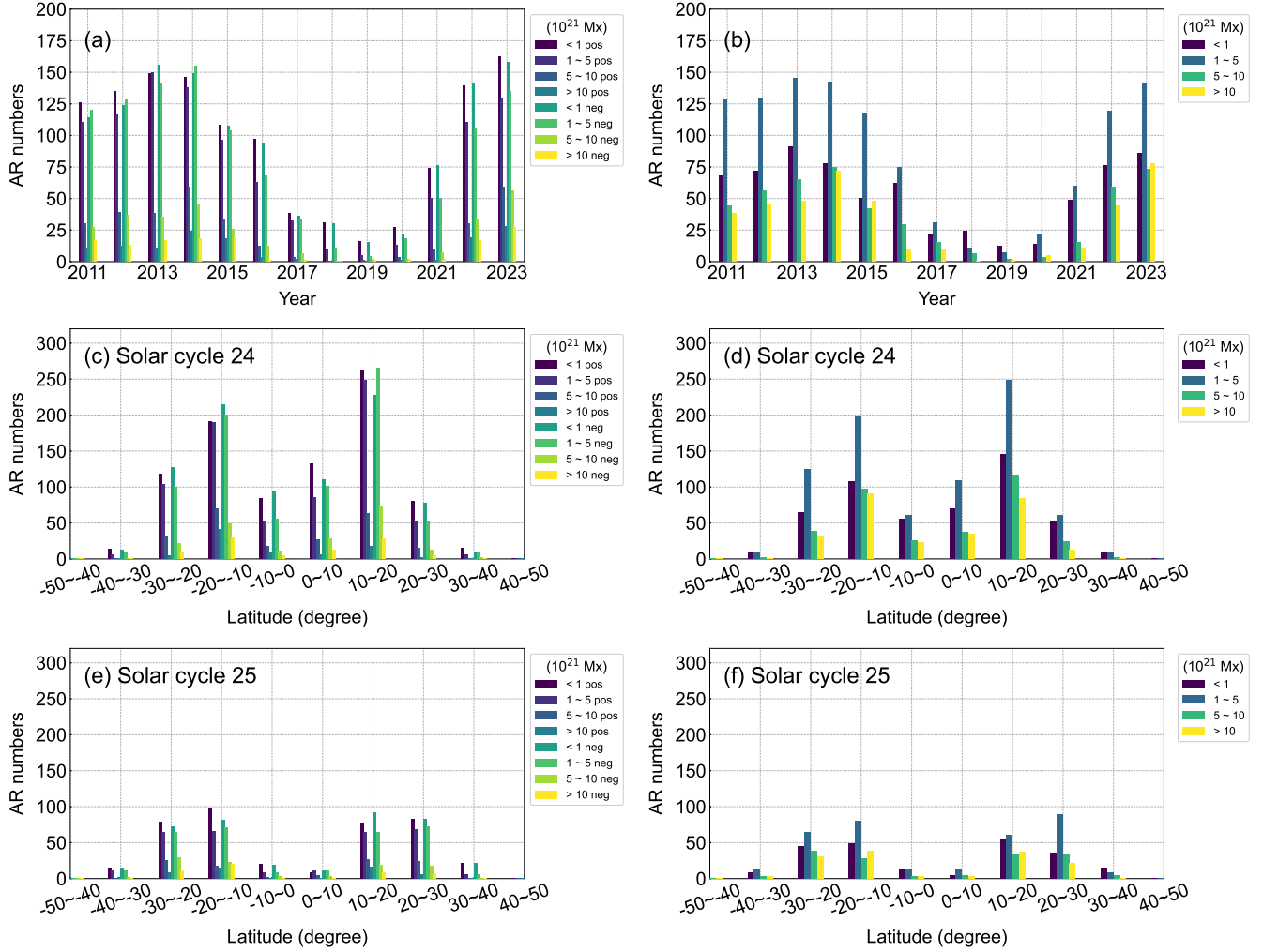


Figure 9. Similar to Figure 7 but for the magnetic flux.

hemisphere within the low-latitude bands of $0^\circ - 30^\circ$. However, this pattern reverses during the rising phase of solar cycle 25, albeit with relatively weaker strength.

3.5. N-S Asymmetry of Active Regions

The N-S asymmetry is a typical characteristic of various solar features, which changes during each solar cycle (Li 2010; Hao et al. 2015; Chowdhury et al. 2019; Veronig et al. 2021; Chandra et al. 2022; Zhukova 2024). The N-S asymmetries of the numbers, areas, and magnetic flux of ARs are also investigated in this work. The N-S asymmetry index is defined as

$$A_{num} = \frac{N_n - N_s}{N_n + N_s}, \quad (3)$$

$$A_{area} \text{ or } A_{f+} \text{ or } A_{f-} \text{ or } A_f = \frac{C_n - C_s}{C_n + C_s}, \quad (4)$$

where N and C denote the numbers and the cumulative area or magnetic flux, while n and s represent the

northern and southern hemispheres, respectively. A_{num} is the N-S asymmetry index for the AR number, A_{area} , A_{f+} , A_{f-} , and A_f are the N-S asymmetry indices for the normalized cumulative positive, negative, and total unsigned magnetic flux, respectively. If the sign of the N-S asymmetry index is positive/negative, it means that the northern/southern hemisphere is dominant.

The yearly asymmetries of the AR number, normalized cumulative area, and fluxes are plotted in Figure 10. The asymmetry indices for the AR number, cumulative area, and total unsigned magnetic flux indicate that the northern hemisphere is dominant for most of the time. The southern hemisphere becomes dominant during 2013–2014 and 2020–2021, and no asymmetry is discernable in 2012, 2015, and 2022. The asymmetry indices for the cumulative positive and negative fluxes exhibit similar trends, with slight deviations. Specifically, the northern hemisphere dominates in the years 2010–2011, 2016–2019, and 2022–2023, while the south-

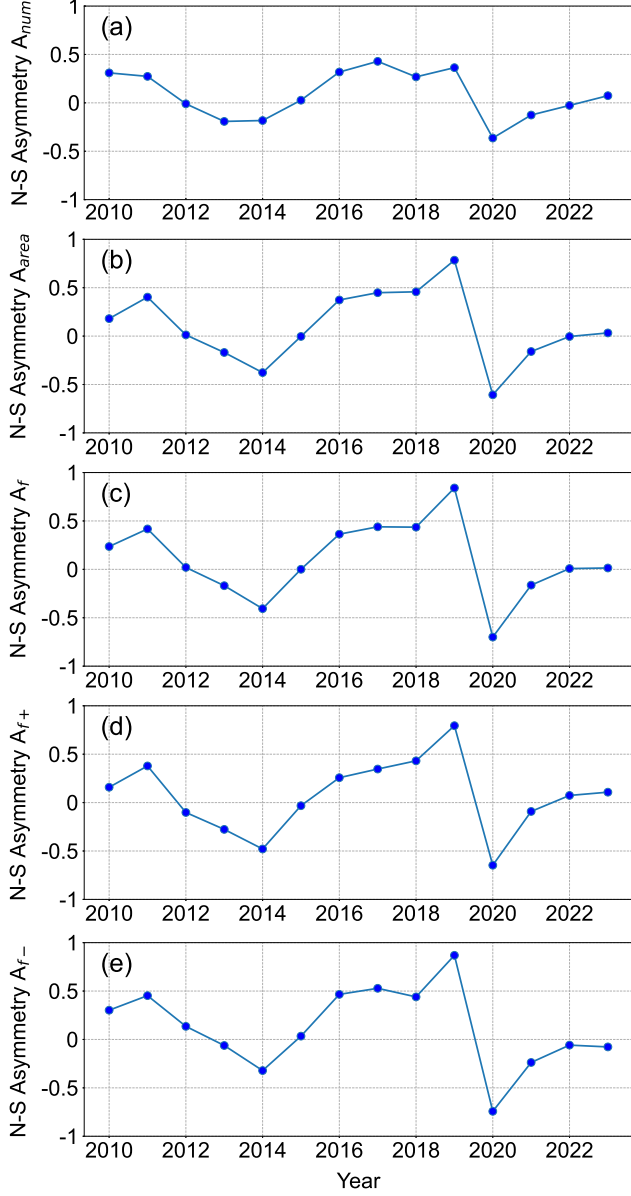


Figure 10. Yearly asymmetries of the number (panel a), cumulative area (panel b), total flux (panel c), cumulative positive (panel d) flux and negative flux (panel e) of all ARs.

ern hemisphere dominates in 2012–2015 and 2020–2021. This suggests that the northern hemisphere is dominant for ARs with both positive and negative magnetic flux during the rising and decay phases of solar cycle 24, whereas the southern hemisphere dominates around the solar maximum. During the rising phase of solar cycle 25, the cumulative positive magnetic flux weakly dominates in the northern hemisphere, and the cumulative negative magnetic flux weakly dominates in the southern hemisphere.

3.6. Distributions of the Tilt Angle of Bipolar Active Regions

The systematic tilt of sunspot groups with respect to the east–west direction was first identified by Hale et al. (1919). The increase of the average tilt angle with latitude—where the leading spot is closer to the equator than the following spot—was termed “Joy’s Law.” The study of active region tilt angles and their long-term variation plays a crucial role in understanding the subsurface dynamics of magnetic flux and the dynamo mechanism (Bhowmik & Nandy 2018). With the high-resolution magnetograms from SOHO/MDI and SDO/HMI, the characteristics of the magnetic tilt angles of bipolar regions have been extensively analyzed (van Driel-Gesztelyi & Green 2015). In our analysis, the magnetic tilt angle of an AR is defined as the angle between the line connecting the centers of the positive and negative polarities and the equator. The sign of the tilt angle is defined as positive if the latitude of the positive polarity region is larger than that of the negative polarity region, and negative if the opposite is true, applicable in both hemispheres.

In accordance with the aforementioned definition, the magnetic tilt angle of each bipolar AR is calculated. The tilt angles are divided into 18 groups at 10-degree intervals, and their yearly histograms are plotted in Figure 11(a). The upward histogram represents the number of ARs with positive tilt angles, while the downward histogram corresponds to the number of ARs with negative tilt angles. It is seen that during solar cycle 24, the number of ARs with positive tilt angles exhibited two peaks: the first in 2011 and the second in 2013. In contrast, the number of ARs with negative tilt angles increased from 2011, peaked in 2014, and then decreased during the declining phase. This number increased again from 2020 to 2023 during the rising phase of solar cycle 25.

Similarly, the histograms of the bipolar AR tilt angles at different latitude bands are shown for solar cycle 24 in Figure 11(b) and the rising phase of solar cycle 25 in Figure 11(c), respectively. Joy’s Law is evident from the two panels as the histograms at larger latitudes have a longer tail, i.e., the tilt angle becomes larger as the latitude increases. During solar cycle 24, there are more ARs with positive tilt angles than those with negative tilt angles in the northern hemisphere, while the opposite is true in the southern hemisphere. This pattern reversed during the rising phase of solar cycle 25.

To clearly illustrate the evolution of the tilt angle distribution, we plot the tilt angles at different latitude bands with vectors, which are overlaid on the butterfly diagram of ARs, as illustrated in Figure 12. The red and black dots represent the positive and negative

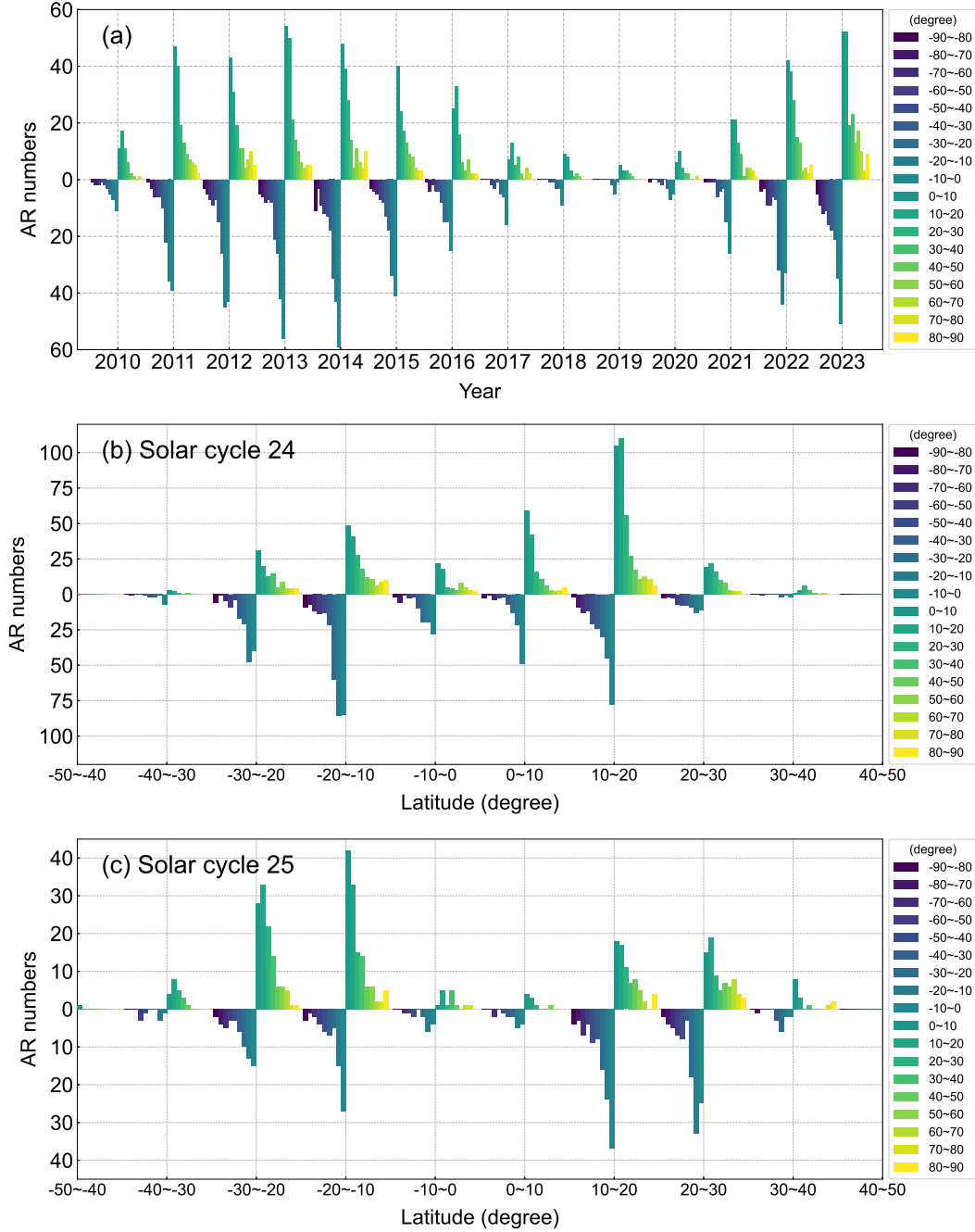


Figure 11. Distribution of the tilt angle of magnetic dipole in ARs with respect to years and latitude from 2010 to 2023. (a) Distribution of the tilt angle of magnetic dipole in ARs with respect to years. (b) Distribution of the tilt angle of magnetic dipole in ARs with respect to latitude during solar cycle 24. Panel (c) is similar to (b) but for those during the rising phase of solar cycle 25.

polarities of bipolar ARs, respectively. Each vector indicates the average tilt angle at specific latitude and year. It is noted that the averaged tilt angle is so small that the difference would be indistinguishable. To make the difference clearer, the orientation of the vectors is amplified by 8 times. From the plot, a discernible pattern consistent with Joy's law is evident for ARs at latitudes

lower than 30° , where the leading polarities are closer to the equator, while the following polarities are farther from the equator. It is also evident that the tilt angles of ARs at relatively high latitudes exhibit a marked increase, resulting in a significant disparity compared to those at lower latitudes. Notably, an anti-Joy's law pattern is observed in ARs in the southern hemisphere

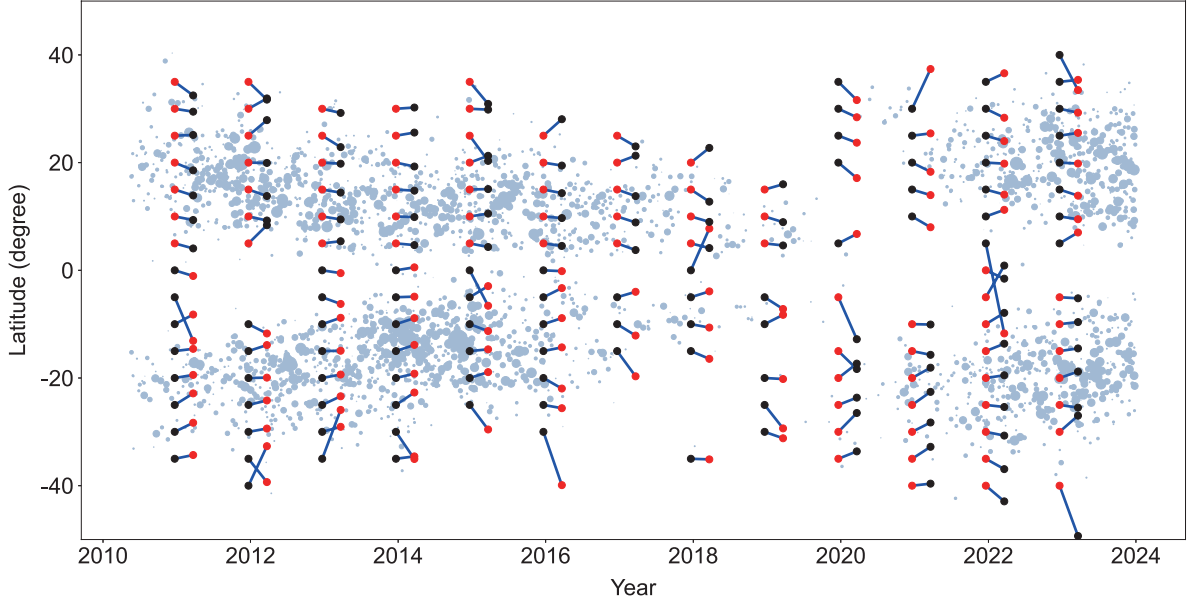


Figure 12. Distribution of average relative latitude of magnetic bipolar ARs with respect to years. The butterfly diagram is represented by light blue dots. The average latitudes of positive and negative polarities are indicated by red and black dots, respectively. The lines connecting the bipoles illustrate the tilt angle at specific latitude and time. To present the difference in tilt angles more clearly, they are multiplied by a factor of 8 to enlarge the angle differences.

during the declining phase of solar cycle 24, spanning from 2017 to 2019.

As indicated by Joy’s law, the tilt angle increases with latitude. This phenomenon has been confirmed by numerous authors through various observations (van Driel-Gesztelyi & Green 2015). We calculate the latitudinal variation of the average tilt angle in different latitude bands at 5° intervals in both hemispheres during solar cycle 24 and the rising phase of solar cycle 25, respectively. The results are shown in Figures 13 and 14. Panels (a)–(d) in Figure 13 illustrate the latitudinal distribution of the average tilt angles of ARs in the area ranges <50 millionths, 50 – 250 millionths, 250 – 500 millionths, and >500 millionths during solar cycle 24, respectively. The blue vertical lines represent the cumulative magnetic flux in each 5° interval. ARs with areas smaller than 50 millionths demonstrate random tilt variations. By considering the cumulative magnetic flux across different latitudes, it is found that the average tilt angle in the latitude bands 5° to 15° in the northern hemisphere, with relatively large cumulative magnetic flux, follows Joy’s law. A similar variation is observed for ARs in the area range of 50 – 250 millionths, but in the southern hemisphere. However, the average tilt angle in the latitude bands 5° to 15° in the northern hemisphere, with relatively large cumulative magnetic flux, deviates from Joy’s law. An opposite tilt angle

pattern is observed for ARs in the area range of 250 – 500 millionths. The average tilt angle of ARs with areas exceeding 500 millionths in the latitude bands 5° to 15° adheres to Joy’s law. ARs at relatively low or high latitudes exhibit random average tilt angles with relatively low cumulative magnetic flux across all area ranges.

As shown in panels (a) and (b) of Figure 14, during the rising phase of solar cycle 25, the average tilt angles of ARs with areas smaller than 250 millionths show random variations, except for those in the area range of 50 – 250 millionths, which follow Joy’s law from 5° to 25° . The average tilt angle of ARs in the area range of 250 – 500 millionths decreases with increasing latitude in the latitude band 5° – 20° , violating Joy’s law. However, larger ARs with areas exceeding 500 millionths follow Joy’s law in the latitude band 5° – 20° .

The error bars indicating 95% confidence for the average tilt angle at different latitudes are plotted in Figures 13 and 14. In both solar cycle 24 and the rising phase of solar cycle 25, the distribution of error bars across different area ranges exhibits a similar pattern: for ARs with areas smaller than 500 millionths, the error bars are very large, often exceeding the range of the average tilt angle distribution at all considered latitudes. In contrast, for ARs with areas larger than 500 millionths, the error bars are relatively smaller, provid-

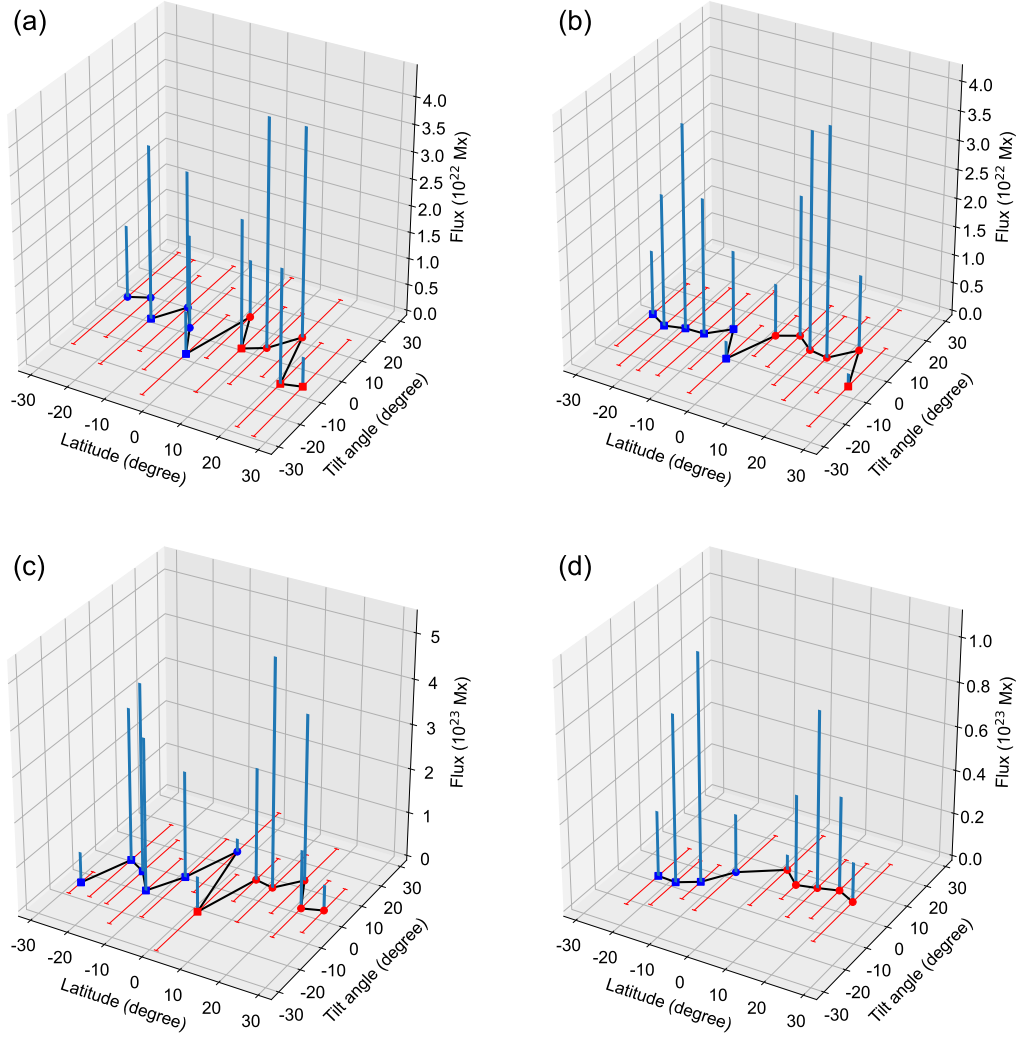


Figure 13. Latitudinal distribution of average ARs tilt angles in every 5° interval during solar cycle 24. Panels (a)–(d) are the latitudinal distribution of average ARs tilt angles in every 5° interval within the area ranges <50 millionth, 50 – 250 millionth, 250 – 500 millionth, and >500 millionth, respectively. The error bars represent the 95% confidence interval, whilst the circles and square symbolize the signs of average title angles. The color red and blue are used to indicate the tilt angle in the northern and southern hemispheres, respectively. The blue vertical lines represent the cumulative magnetic flux in every 5° interval.

ing greater confidence that Joy’s law is satisfied in the latitude range of $5^\circ - 20^\circ$.

3.7. Distributions of the Regular and Anti-Hale Active Regions

Hale’s law states that the leading polarity of bipolar ARs is negative in the northern hemisphere and positive in the southern hemisphere during even solar cycles, and vice versa during odd solar cycles (Hale et al. 1919). The detected ARs in this study conform to Hale’s law, as shown in Figure 12. Specifically, the leading polarities are negative in the northern hemisphere and positive in the southern hemisphere during solar cycle 24, while the opposite is observed in the rising phase of

solar cycle 25. However, it is important to note that some ARs exhibit distinct morphological characteristics and violate Hale’s law. These so-called “anti-Hale” ARs have been reported by numerous authors (Stenflo & Kosovichev 2012; van Driel-Gesztelyi & Green 2015; Li 2018; Zhukova et al. 2020; Muñoz-Jaramillo et al. 2021; Zhukova 2024). In this study, we identify both regular and anti-Hale ARs among the detected ARs and plot the results in Figure 15. The red and blue dots represent the latitudes of regular and anti-Hale ARs, respectively. As shown in Figure 15(a), anti-Hale ARs are distributed across all latitudes below 40° and throughout the entirety of solar cycle 24. A notable increase in the number of anti-Hale ARs was observed in 2014, par-

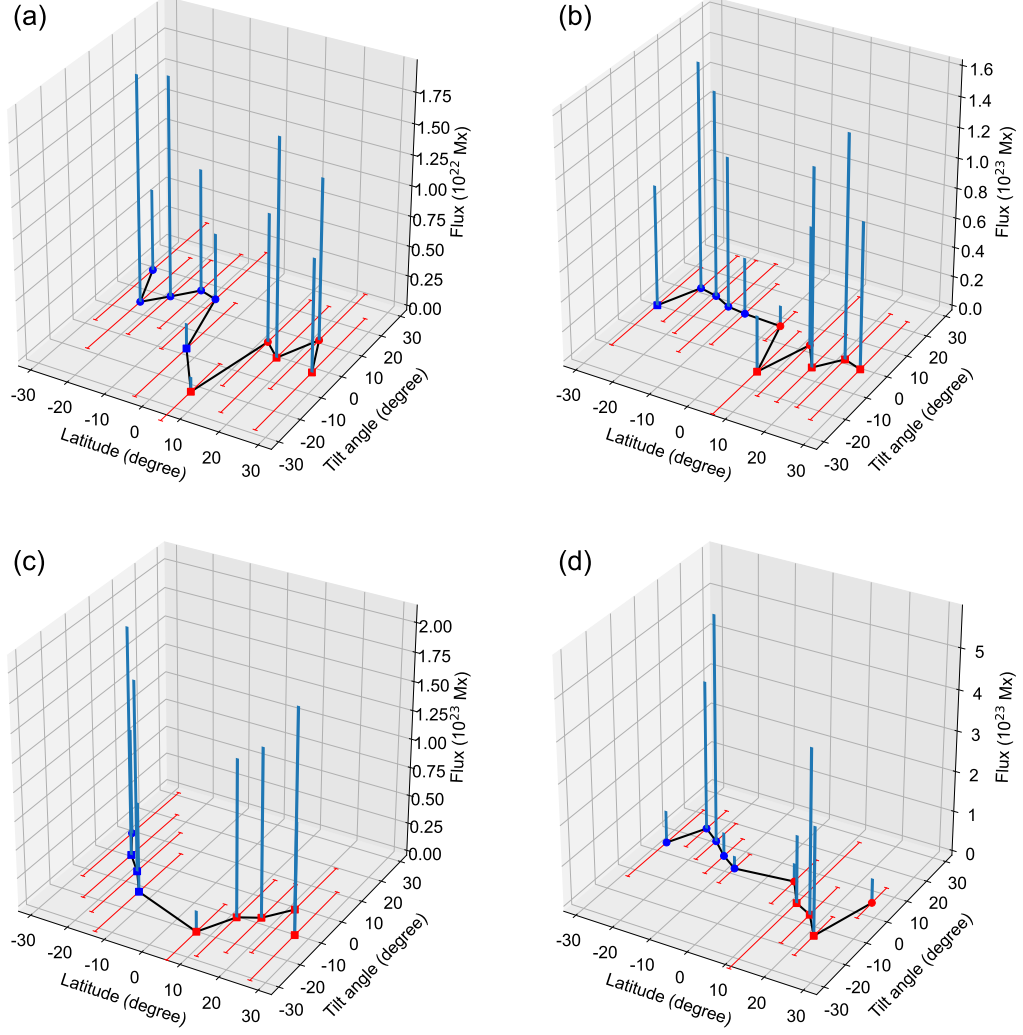


Figure 14. Similar to Figure 13 but for the rising phase of solar cycle 25.

ticularly in the southern hemisphere, around the time of solar maximum. The latitudinal distribution of both regular and anti-Hale ARs exhibits a bimodal pattern, with peaks between 10° and 20° in both hemispheres. A similar trend is observed during the rising phase of solar cycle 25. It is revealed that the proportion of ARs violating Hale’s law is 13% and 16% in solar cycles 24 and 25, respectively.

4. DISCUSSION AND CONCLUSIONS

In this study, an automated detection method for solar ARs in LOS full-disk magnetograms was developed. The method, based on the DBSCAN clustering algorithm, was named DBSCAN²-based Solar Active Region Detection (DSARD). It is an unsupervised machine learning method that does not require pre-labeled active regions as training samples. The validity of the method was established by using the NOAA catalogue

as the reference standard. The average true positive rate (R_{TP}) and false positive rate (R_{FP}) of the method were determined to be 91.8% and 7.2%, respectively. Subsequently, LOS full-disk magnetograms observed by SDO/HMI from May 2010 to December 2023 were processed using this method to investigate the long-term variation of ARs. To avoid multiple detections and to minimize the projection effects, we examined the ARs only within a longitudinal range of $\pm 6^\circ$ from the central meridian of the solar disk. A total of 1991 ARs were detected within this range during solar cycle 24, and 872 ARs were detected during the rising phase of solar cycle 25. The evolution of these active regions presented the typical butterfly diagram. We calculated the drift velocities, the location, area, magnetic flux, and tilt angles of these ARs across different years and latitude bands.

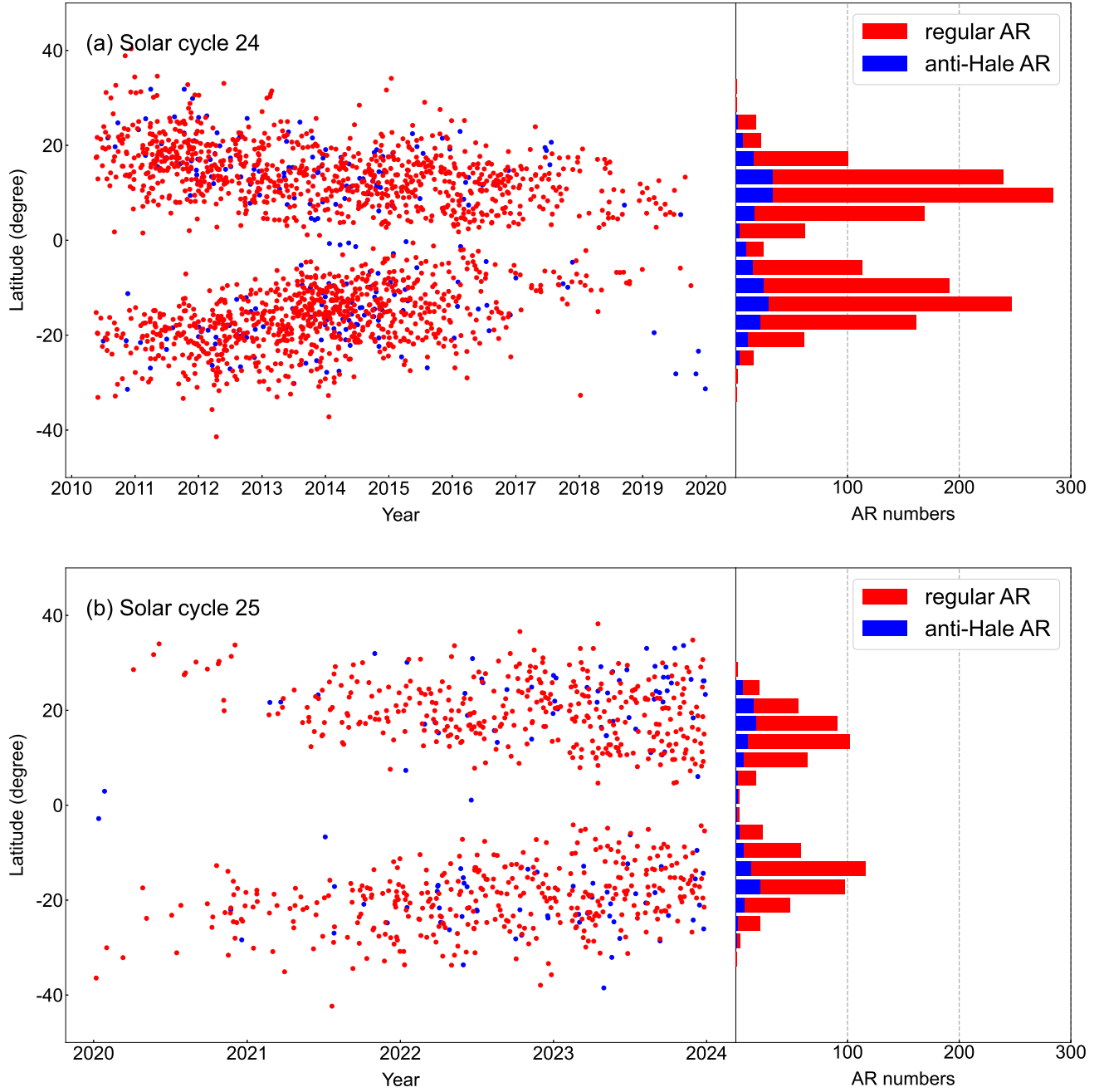


Figure 15. Latitudinal distribution of ARs categorized by whether they violate Hale’s law. (a) Latitudinal distribution of regular and anti-Hale ARs during solar cycle 24, with red and blue dots representing regular and anti-Hale ARs, respectively. The histogram on the right side of the figure presents the distribution of AR numbers for regular and anti-Hale ARs in different latitude bands. Panel (c) is similar to (b) but for those during the rising phase of solar cycle 25.

A strong correlation exists between the yearly variation of AR numbers detected by our method and those recorded in the NOAA catalogue from 2010 to 2023, although our method yields comparatively more ARs. This discrepancy can be attributed to the enhanced sensitivity of our method, which enables detecting relatively small and diffuse ARs. Our analysis revealed

the presence of two peaks: the first occurs in the northern hemisphere around 2012, and the second appears in the southern hemisphere around 2014. This finding aligns with observations of sunspots reported in previous studies (Chowdhury et al. 2019; Veronig et al. 2021; Nandy 2021; Du 2022; Andreeva 2023). During the rising phase of solar cycle 25, the number of ARs increased

more rapidly than during the early phase of solar cycle 24, suggesting that the strength of solar cycle 25 may exceed that of solar cycle 24. The latitudinal distribution of AR numbers exhibits a bimodal pattern, with a majority of ARs concentrated within the latitude band of 10° – 20° and few above 40° .

The latitudinal migration of ARs is illustrated in the butterfly diagrams shown in Figure 6 for solar cycle 24 and the rising phase of solar cycle 25. Relatively large ARs began to appear in the northern hemisphere around 2012, while in the southern hemisphere, they emerged in late 2012 and became particularly large and numerous in 2014 and 2015. This indicates that the ARs in the southern hemisphere were stronger than those in the northern hemisphere during the solar maximum of solar cycle 24. In solar cycle 25, the solar activity ascended more rapidly, and opposite to the previous cycle, ARs are dominant in the southern hemisphere.

To quantitatively analyze the migration of ARs during solar cycle 24 and the rising phase of solar cycle 25, the monthly average latitudes of ARs were calculated, and a cubic polynomial fitting was applied to derive their drift velocities. During solar cycle 24, ARs in the southern hemisphere exhibited relatively stable drift velocities, ranging between $[0.7, 0.8] \text{ m s}^{-1}$. In contrast, the drift velocities of ARs in the northern hemisphere showed more pronounced variations. They initially decreased from 1.5 m s^{-1} to 0.1 m s^{-1} in late 2015, before increasing to 0.8 m s^{-1} by the end of solar cycle 24 in 2019. During the rising phase of solar cycle 25, clear asymmetries in drift velocities were also observed between the two hemispheres. In the northern hemisphere, the drift velocities decelerated towards higher latitudes from 2021 to 2022, followed by an acceleration towards the equator, reaching 1.0 m s^{-1} in 2023. Conversely, the southern hemisphere exhibited a uniform acceleration of drift velocities towards the equator, increasing from 0.5 m s^{-1} to 1.2 m s^{-1} .

Li (2010) analyzed the drift velocities of sunspot groups from 1919 to 1989, covering six solar cycles recorded by the Royal Greenwich Observatory and NOAA. They identified a two-stage migration pattern: sunspot groups initially migrate from latitudes of approximately 28° with a drift velocity of about 1.2 m s^{-1} towards the solar equator from the beginning of a cycle to the cycle maximum. This is followed by a transition to a drift velocity of about 1.0 m s^{-1} at latitudes of approximately 20° towards the solar equator during the remainder of the solar cycle, ultimately reaching a final latitude of about 8° . The results obtained in this study demonstrate that the migration of ARs also exhibited a two-phase pattern within solar cycle 24: from

the beginning to the solar maximum, the drift velocity decelerates, followed by an acceleration during the declining phase. Zhang et al. (2010) applied their automated detection method to identify ARs in synoptic magnetograms constructed by SOHO/MDI from 1996 to 2008 during solar cycle 23. They found a linear drifting mode and derived an average drift velocity of $0.708 \pm 0.015 \text{ m s}^{-1}$. The drift velocities derived from the ARs detected in the present study are consistent with those reported in previous studies, although minor discrepancies may arise due to differences in data sources and time ranges.

The AR areas were divided into four groups based on their sizes: <50 millionths, 50 – 250 millionths, 250 – 500 millionths, and >500 millionths, with the proportions being approximately $1 : 2 : 4 : 2$. The latitudinal distribution of the AR numbers in the four area groups, as well as their cumulative areas during solar cycle 24 and the rising phase of solar cycle 25, exhibited bimodal patterns. Only a few ARs are located in the latitude band above 30° in both hemispheres. The peaks of AR numbers occurred in the latitude band of 10° – 20° , with the exception of ARs in the area range of 50 – 250 millionths in the northern hemisphere and those in the range of 250 – 500 millionths in the southern hemisphere. Notably, the AR numbers in the latitude band of 30° – 40° during the rising phase of solar cycle 25 were comparatively larger than those in solar cycle 24. This suggests that ARs were distributed over a wider range of latitudes in solar cycle 25.

The latitudinal distribution of the AR numbers in the four area groups and their cumulative area in solar cycle 24 and the rising phase of solar cycle 25 manifested as bimodal distributions, with a few ARs occupying the latitude band over 30° in both hemispheres. The peaks of AR numbers appeared in the latitude band 10° – 20° , with the exception of ARs with area within 50 – 250 millionth in the northern hemisphere and that with area within 250 – 500 millionth in the southern hemisphere. It is noteworthy that the AR numbers in the latitude band 30° – 40° during the rising phase of solar cycle 25 were comparatively larger than those in solar cycle 24. This observation suggests that the ARs were distributed over a wider range of latitudes in solar cycle 25.

The magnetic flux was calculated by summing the magnetic fluxes in each AR, yielding the positive, negative, and total unsigned fluxes for each individual AR, respectively. Similar to AR areas, the AR magnetic fluxes were also divided into four groups based on their intensities: $< 10^{21} \text{ Mx}$, 1 – $5 \times 10^{21} \text{ Mx}$, 5 – $10 \times 10^{21} \text{ Mx}$, and $> 10^{22} \text{ Mx}$. It was observed that the magnetic flux exhibits a yearly and latitudinal distribution consistent

with that of the AR areas, indicating a correlation between the two variables. The net magnetic flux was found to be negative from 2010 to 2013, becoming positive with relatively large values in 2014, concurrent with the polar reversal in solar cycle 24. Subsequently, the net magnetic flux remained near zero during the decay phase of solar cycle 24. During the rising phase of solar cycle 25, the positive and negative magnetic fluxes exhibited rapid growth. However, the net magnetic flux remained approximately zero. The latitudinal distribution of the cumulative net magnetic flux of ARs revealed that, in solar cycle 24, the net magnetic flux is positive in the southern hemisphere and negative in the northern hemisphere within the relatively low latitude bands of 0° to 30° . Conversely, during the rising phase of solar cycle 25, the net magnetic flux exhibits a reverse trend, albeit with comparatively weaker strength.

The asymmetry indices of the AR number, cumulative area, and total unsigned magnetic flux revealed that the northern hemisphere was dominant across all types of ARs, with the exception of the periods 2013–2014 and 2020–2021. Notable asymmetry indices were also observed in 2012, 2015, and 2022. The asymmetry indices for cumulative positive and negative magnetic fluxes exhibited similar trends, albeit with some exceptions. These indices suggest that the northern hemisphere dominated in ARs with both positive and negative magnetic flux during the rising and decay phases of solar cycle 24, while the southern hemisphere becomes dominant around the solar maximum. During the rising phase of solar cycle 25, northern hemisphere was dominant for the cumulative positive magnetic flux, whereas the opposite is true for cumulative negative magnetic flux. However, these asymmetry indices are not pronounced.

The magnetic tilt angle of each bipolar AR was calculated, and its distribution was analyzed with respect to years and latitude bands. The proportion of ARs with large tilt angles increased with latitude, a phenomenon consistent with Joy’s law. To provide a clear and intuitive representation of the tilt angle distribution, the mean tilt angles of bipolar ARs in different latitude bands were plotted and overlaid on the butterfly diagram of ARs. The long-term variation of tilt angles is presented in Figure 12. A clear pattern consistent with Joy’s law is evident for the ARs at latitudes lower than 30° . At higher latitudes, the tilt angles of ARs increase, accompanied by a widening disparity compared to those at lower latitudes. Notably, an anti-Joy’s law pattern was observed in the southern hemisphere as solar cycle 24 approached minimum, spanning from 2017 to 2019.

Tlatov et al. (2013) discovered that bipolar ARs with areas exceeding 300 millionths of the solar hemisphere (MHS) generally follow Joy’s law, whereas those with areas in the range of 50 to 300 MHS exhibit an inverse pattern to Joy’s law. In this work, a similar tendency to follow Joy’s law was identified in relatively large ARs with areas over 500 millionths, particularly within the high cumulative magnetic flux latitude band of 5° to 20° . In contrast, smaller bipolar ARs displayed random variations in their average tilt angles. The observed randomness in the results can be attributed to the wide error bars associated with small bipolar ARs. These large error bars are likely due to significant fluctuations at all scales (Stenflo & Kosovichev 2012). Furthermore, Tlatov et al. (2013) reported a bending of Joy’s law at 30° for bipolar ARs with areas exceeding 300 MHS, while those with areas between 50 and 300 MHS exhibited bending at latitudes ranging from 30° to 40° . Regarding the findings of this study, it is noteworthy that, if the influence of error bars is disregarded and the focus is placed solely on the average tilt angle, deviations from Joy’s law are observed. Specifically, large ARs (with areas over 500 millionths) show deviations at 20° latitude, while relatively small ARs (with areas between 50 and 250 millionths) in the southern hemisphere exhibit deviations at 25° latitude.

The detected ARs exhibited strong adherence to Hale’s law, with leading polarities being negative in the northern hemisphere and positive in the southern hemisphere during solar cycle 24. This pattern reversed during the rising phase of solar cycle 25. However, it is important to note that a subset of ARs violated Hale’s law, accounting for 13% and 16% of all ARs during solar cycle 24 and the rising phase of solar cycle 25, respectively. Zhukova et al. (2020) reported that approximately 3.0% of all studied ARs were anti-Hale regions, based on their selection criteria. In contrast, recent studies found higher percentages of anti-Hale ARs, ranging from 4% to 8% (Khlystova & Sokoloff 2009; McClintock et al. 2014; Li 2018). Notably, the findings of our study, reaching 13% and 16%, align with these recent results. Zhukova et al. (2020) proposed several reasons that may explain these observations. First, small, short-lived ARs are more likely to exhibit anti-Hale patterns. Second, the enhanced sensitivity of modern instruments enables the detection of weaker ARs. Third, identifying anti-Hale ARs within activity complexes remains a significant challenge. Additionally, the high sensitivity of the DSARD method in detecting ARs may also contribute to the observed results. The distribution of anti-Hale ARs spans all latitudes below 40° and extends throughout the entirety of solar cy-

cle 24. A notable increase in the number of anti-Hale ARs was observed in 2014, particularly in the southern hemisphere, coinciding with the solar maximum. The latitudinal distributions of both regular and anti-Hale ARs exhibit a bimodal pattern, with peaks occurring in the latitude bands of 10° to 20° in both hemispheres. It is worth noting that similar trends were observed during the rising phase of solar cycle 25.

In summary, an automated detection method for solar ARs in LOS full-disk magnetograms was developed. This method is based on the DBSCAN clustering algorithm. The validity and accuracy of the method was demonstrated through comparison with the NOAA catalogue. The method was then applied to process the LOS full-disk magnetograms observed by SDO/HMI from May 2010 to December 2023 to obtain the long-term variation of ARs. The key results are summarized as follows:

1. The total number of ARs peaked around 2014, with the maximum number occurring in 2012 in the northern hemisphere and in 2014 in the southern hemisphere. The total area and total magnetic flux reached their maxima in late 2014. Approximately 99% of ARs were located at latitudes below 40° , with their latitudinal distribution peaking in the $[10^\circ, 20^\circ]$ band in both hemispheres.
2. The latitudinal distributions of the total number, area, and magnetic flux of ARs were predominantly bimodal, with peaks in the $[10^\circ, 20^\circ]$ latitude band in both hemispheres.
3. The latitudinal migration of ARs during solar cycle 24 exhibited two distinct phases: from the beginning of the cycle to the year following the solar maximum, the drift velocity decelerated, then ac-

celerated during the declining phase. The drift velocities ranged between $[0.1, 1.6] \text{ m s}^{-1}$. Variations in drift velocities were significantly more pronounced in the northern hemisphere compared to the southern hemisphere, both during solar cycle 24 and the rising phase of solar cycle 25.

4. The asymmetry indices of the AR number, cumulative area, and total unsigned magnetic flux all indicate that the northern hemisphere dominated in terms of AR activity, except during the periods 2013–2014 and 2020–2021.
5. Bipolar ARs with latitudes below 30° follow Joy’s law. However, the tilt angles of ARs at latitudes above 30° exhibit no clear pattern.
6. ARs with areas smaller than 500 millionths showed random variations in the average tilt angles. In contrast, ARs with areas larger than 500 millionths and high cumulative magnetic flux in the 5° to 20° latitude band followed Joy’s law, with a bending of Joy’s law observed above the latitude 20° .
7. Approximately 13% and 16% of all ARs violated Hale’s law during solar cycle 24 and the rising phase of solar cycle 25, respectively.

We are grateful to the SDO teams for providing the observational data. The present work was supported by the National Key Research and Development Program of China (2020YFC2201200), NSFC under grants 12173019 and 12127901, and the Young Data Scientist Program of the China National Astronomical Data Center, as well as the AI & AI for Science Project of Nanjing University.

REFERENCES

- Andreeva, O. 2023, *Advances in Space Research*, 71, 1915, doi: [10.1016/j.asr.2022.07.043](https://doi.org/10.1016/j.asr.2022.07.043)
- Barnes, G., Leka, K. D., Schrijver, C. J., et al. 2016, *ApJ*, 829, 89, doi: [10.3847/0004-637X/829/2/89](https://doi.org/10.3847/0004-637X/829/2/89)
- Barra, V., Delouille, V., & Hochedez, J.-F. 2008, *Advances in Space Research*, 42, 917, doi: <https://doi.org/10.1016/j.asr.2007.10.021>
- Barra, V., Delouille, V., Kretschmar, M., & Hochedez, J. F. 2009, *A&A*, 505, 361, doi: [10.1051/0004-6361/200811416](https://doi.org/10.1051/0004-6361/200811416)
- Benkhalil, A., Zharkova, V. V., Zharkov, S., & Ipson, S. 2006, *SoPh*, 235, 87, doi: [10.1007/s11207-006-0023-7](https://doi.org/10.1007/s11207-006-0023-7)
- Bhowmik, P., & Nandy, D. 2018, *Nature Communications*, 9, 5209, doi: [10.1038/s41467-018-07690-0](https://doi.org/10.1038/s41467-018-07690-0)
- Bobra, M. G., & Couvidat, S. 2015, *ApJ*, 798, 135, doi: [10.1088/0004-637X/798/2/135](https://doi.org/10.1088/0004-637X/798/2/135)
- Bobra, M. G., Sun, X., Hoeksema, J. T., et al. 2014, *SoPh*, 289, 3549, doi: [10.1007/s11207-014-0529-3](https://doi.org/10.1007/s11207-014-0529-3)
- Caballero, C., & Aranda, M. C. 2013, *SoPh*, 283, 691, doi: [10.1007/s11207-013-0239-2](https://doi.org/10.1007/s11207-013-0239-2)
- Chandra, Y., Pande, B., Mathpal, M. C., & Pande, S. 2022, *Astrophysics*, 65, 404, doi: [10.1007/s10511-022-09749-9](https://doi.org/10.1007/s10511-022-09749-9)
- Chen, P. F. 2011, *Living Reviews in Solar Physics*, 8, 1, doi: [10.12942/lrsp-2011-1](https://doi.org/10.12942/lrsp-2011-1)

- Choudhuri, A. R. 2021, *Science China Physics, Mechanics, and Astronomy*, 64, 239601, doi: [10.1007/s11433-020-1628-1](https://doi.org/10.1007/s11433-020-1628-1)
- Chowdhury, P., Kilcik, A., Yurchyshyn, V., Obridko, V. N., & Rozelot, J. P. 2019, *SoPh*, 294, 142, doi: [10.1007/s11207-019-1530-7](https://doi.org/10.1007/s11207-019-1530-7)
- Colak, T., & Qahwaji, R. 2008, *SoPh*, 248, 277, doi: [10.1007/s11207-007-9094-3](https://doi.org/10.1007/s11207-007-9094-3)
- Cui, Y.-m., Liu, S.-q., & Shi, L.-q. 2021, *ChA&A*, 45, 458, doi: [10.1016/j.chinastron.2021.11.002](https://doi.org/10.1016/j.chinastron.2021.11.002)
- Du, Z. 2022, *SoPh*, 297, 70, doi: [10.1007/s11207-022-02005-5](https://doi.org/10.1007/s11207-022-02005-5)
- Guo, X., Yang, Y., Feng, S., et al. 2022, *SoPh*, 297, 104, doi: [10.1007/s11207-022-02019-z](https://doi.org/10.1007/s11207-022-02019-z)
- Hale, G. E., Ellerman, F., Nicholson, S. B., & Joy, A. H. 1919, *ApJ*, 49, 153, doi: [10.1086/142452](https://doi.org/10.1086/142452)
- Hao, Q., Fang, C., Cao, W., & Chen, P. F. 2015, *ApJS*, 221, 33, doi: [10.1088/0067-0049/221/2/33](https://doi.org/10.1088/0067-0049/221/2/33)
- Hathaway, D. H. 2010, *Living Reviews in Solar Physics*, 7, 1, doi: [10.12942/lrsp-2010-1](https://doi.org/10.12942/lrsp-2010-1)
- Higgins, P., Gallagher, P., McAteer, R., & Bloomfield, D. 2011, *Advances in Space Research*, 47, 2105, doi: <https://doi.org/10.1016/j.asr.2010.06.024>
- Jiang, B., Liu, L., Zheng, S., et al. 2022, *ChA&A*, 46, 264, doi: [10.1016/j.chinastron.2022.09.007](https://doi.org/10.1016/j.chinastron.2022.09.007)
- Khlystova, A. I., & Sokoloff, D. D. 2009, *Astronomy Reports*, 53, 281, doi: [10.1134/S106377290903010X](https://doi.org/10.1134/S106377290903010X)
- Lemen, J. R., Title, A. M., Akin, D. J., et al. 2012, *SoPh*, 275, 17, doi: [10.1007/s11207-011-9776-8](https://doi.org/10.1007/s11207-011-9776-8)
- Li, J. 2018, *ApJ*, 867, 89, doi: [10.3847/1538-4357/aae31a](https://doi.org/10.3847/1538-4357/aae31a)
- Li, K. J. 2010, *MNRAS*, 405, 1040, doi: [10.1111/j.1365-2966.2010.16508.x](https://doi.org/10.1111/j.1365-2966.2010.16508.x)
- Liu, D., Song, W., Lin, G., & Wang, H. 2021, *SoPh*, 296, 176, doi: [10.1007/s11207-021-01920-3](https://doi.org/10.1007/s11207-021-01920-3)
- Liu, H., Xu, Y., Wang, J., et al. 2020, *ApJ*, 894, 70, doi: [10.3847/1538-4357/ab8818](https://doi.org/10.3847/1538-4357/ab8818)
- Maunder, E. W. 1904, *MNRAS*, 64, 747, doi: [10.1093/mnras/64.8.747](https://doi.org/10.1093/mnras/64.8.747)
- McAteer, R. T. J., Gallagher, P. T., & Ireland, J. 2005a, *ApJ*, 631, 628, doi: [10.1086/432412](https://doi.org/10.1086/432412)
- McAteer, R. T. J., Gallagher, P. T., Ireland, J., & Young, C. A. 2005b, *SoPh*, 228, 55, doi: [10.1007/s11207-005-4075-x](https://doi.org/10.1007/s11207-005-4075-x)
- McClintock, B. H., Norton, A. A., & Li, J. 2014, *ApJ*, 797, 130, doi: [10.1088/0004-637X/797/2/130](https://doi.org/10.1088/0004-637X/797/2/130)
- Muñoz-Jaramillo, A., Navarrete, B., & Campusano, L. E. 2021, *ApJ*, 920, 31, doi: [10.3847/1538-4357/ac133b](https://doi.org/10.3847/1538-4357/ac133b)
- Nandy, D. 2021, *SoPh*, 296, 54, doi: [10.1007/s11207-021-01797-2](https://doi.org/10.1007/s11207-021-01797-2)
- Quan, L., Xu, L., Li, L., Wang, H., & Huang, X. 2021, *Electronics*, 10, doi: [10.3390/electronics10182284](https://doi.org/10.3390/electronics10182284)
- Ravindra, B., Chowdhury, P., Ray, P. C., & Pichamani, K. 2022, *ApJ*, 940, 43, doi: [10.3847/1538-4357/ac98b3](https://doi.org/10.3847/1538-4357/ac98b3)
- Ronneberger, O., Fischer, P., & Brox, T. 2015, *arXiv e-prints*, arXiv:1505.04597, doi: [10.48550/arXiv.1505.04597](https://doi.org/10.48550/arXiv.1505.04597)
- Scherrer, P. H., Bogart, R. S., Bush, R. I., et al. 1995, *SoPh*, 162, 129, doi: [10.1007/BF00733429](https://doi.org/10.1007/BF00733429)
- Schou, J., Scherrer, P. H., Bush, R. I., et al. 2012, *SoPh*, 275, 229, doi: [10.1007/s11207-011-9842-2](https://doi.org/10.1007/s11207-011-9842-2)
- Schrijver, C. J. 2007, *ApJL*, 655, L117, doi: [10.1086/511857](https://doi.org/10.1086/511857)
- Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. 2017, *ACM Trans. Database Syst.*, 42, doi: [10.1145/3068335](https://doi.org/10.1145/3068335)
- Solanki, S. K. 2003, *A&A Rv*, 11, 153, doi: [10.1007/s00159-003-0018-4](https://doi.org/10.1007/s00159-003-0018-4)
- Stenflo, J. O., & Kosovichev, A. G. 2012, *ApJ*, 745, 129, doi: [10.1088/0004-637X/745/2/129](https://doi.org/10.1088/0004-637X/745/2/129)
- Sun, Z., Bobra, M. G., Wang, X., et al. 2022, *ApJ*, 931, 163, doi: [10.3847/1538-4357/ac64a6](https://doi.org/10.3847/1538-4357/ac64a6)
- Tlatov, A., Illarionov, E., Sokoloff, D., & Pipin, V. 2013, *MNRAS*, 432, 2975, doi: [10.1093/mnras/stt659](https://doi.org/10.1093/mnras/stt659)
- Turmon, M., Jones, H. P., Malanushenko, O. V., & Pap, J. M. 2010, *SoPh*, 262, 277, doi: [10.1007/s11207-009-9490-y](https://doi.org/10.1007/s11207-009-9490-y)
- Turmon, M., Pap, J. M., & Mukhtar, S. 2002, *ApJ*, 568, 396, doi: [10.1086/338681](https://doi.org/10.1086/338681)
- van Driel-Gesztelyi, L., & Green, L. M. 2015, *Living Reviews in Solar Physics*, 12, 1, doi: [10.1007/lrsp-2015-1](https://doi.org/10.1007/lrsp-2015-1)
- Verbeeck, C., Higgins, P. A., Colak, T., et al. 2013, *SoPh*, 283, 67, doi: [10.1007/s11207-011-9859-6](https://doi.org/10.1007/s11207-011-9859-6)
- Veronig, A. M., Jain, S., Podladchikova, T., Pötzi, W., & Clette, F. 2021, *A&A*, 652, A56, doi: [10.1051/0004-6361/202141195](https://doi.org/10.1051/0004-6361/202141195)
- Zhang, H., Li, Q., Yang, Y., et al. 2022, *ApJS*, 263, 28, doi: [10.3847/1538-4365/ac9b17](https://doi.org/10.3847/1538-4365/ac9b17)
- Zhang, J., Wang, Y., & Liu, Y. 2010, *ApJ*, 723, 1006, doi: [10.1088/0004-637X/723/2/1006](https://doi.org/10.1088/0004-637X/723/2/1006)
- Zhang, T., Hao, Q., & Chen, P. F. 2024, *ApJS*, 272, 5, doi: [10.3847/1538-4365/ad3039](https://doi.org/10.3847/1538-4365/ad3039)
- Zharkov, S., Zharkova, V. V., & Ipson, S. S. 2005, *SoPh*, 228, 377, doi: [10.1007/s11207-005-5005-7](https://doi.org/10.1007/s11207-005-5005-7)
- Zhu, G., Lin, G., Wang, D., Liu, S., & Yang, X. 2019, *SoPh*, 294, 117, doi: [10.1007/s11207-019-1517-4](https://doi.org/10.1007/s11207-019-1517-4)
- Zhukova, A. 2024, *MNRAS*, 532, 2032, doi: [10.1093/mnras/stae1604](https://doi.org/10.1093/mnras/stae1604)
- Zhukova, A., Khlystova, A., Abramenko, V., & Sokoloff, D. 2020, *SoPh*, 295, 165, doi: [10.1007/s11207-020-01734-9](https://doi.org/10.1007/s11207-020-01734-9)