Publications of the Astronomical Society of the Pacific, 130:104503 (14pp), 2018 October https://doi.org/10.1088/1538-3873/aadbfa

Automatic Detection of Sunspots on Full-disk Solar Images Using the Simulated Annealing Genetic Method

Yunfei Yang[1,2], Hongjuan Yang[1], Xianyong Bai[2], Huituan Zhou[3], Song Feng[1], and Bo Liang[1]
[1] Faculty of Information Engineering and Automation/Yunnan Key Laboratory of Computer Technology Application, Kunming University of Science and Technology, Kunming 650500, Yunnan, People's Republic of China; yangyf@escience.cn
[2] CAS Key Laboratory of Solar Activity, National Astronomical Observatories, Beijing 100012, People's Republic of China
[3] China Telecom Corporation Limited Yunnan Branch, Kunming 650200, Yunnan, People's Republic of China
Received 2018 July 11; accepted 2018 August 20; published 2018 September 7

Abstract

Sunspots with strong magnetic fields are the most important manifestations of solar activity, appearing as dark features in the photosphere observed in continuum images. We proposed an artificial intelligence technology called the simulated annealing genetic (SAG) method, which combined the genetic algorithm and simulated annealing algorithm to self-adaptively derive dual thresholds for detecting the umbra and penumbra of sunspots simultaneously. Full-disk continuum intensity images obtained from *Solar Dynamics Observatory*/Helioseismic Magnetic Imager (HMI) at a cadence of four hours from 2010 May to 2016 December were used. The detection results showed that the dual thresholds derived by the SAG method have outstanding performance in segmenting the umbra and penumbra from the photosphere with a satisfactory robustness efficiently. The boundaries of the umbra and penumbra were finely delineated, even for sunspots at the extreme solar limb. The total sunspot areas, umbral areas, and penumbral areas match very well with the data reported from HMI Debrecen Data (HMIDD), with the correlation coefficients reaching 0.99, 0.99, and 0.95, respectively. The mean ratios of umbra to sunspot areas per year ranged from 0.159 to 0.233. The ratios decreased with an increase in solar activity, which implies that the ratio was related to the solar activity level.

Key words: methods: data analysis – Sun: activity – sunspots – techniques: image processing

Online material: color figures

1. Introduction

Sunspots are visibly darker than the photosphere in white-light continuum images. Within sunspots, the darker cores are called the umbra and the peripheral, relatively bright regions are called the penumbra. They have different mechanisms for formation and maintenance (Li et al. 2018). Sunspots with strong magnetic fields are the most obvious phenomenon in the solar photosphere and are important manifestations of solar activities (Yan & Qu 2007; Yan et al. 2016, 2018a, 2018b). It is, therefore, important to detect the umbra and penumbra of sunspots by a robust and reliable automated detection method with high precision.

Many automated detection methods for sunspots have been proposed in recent years owing to the large number of high-resolution images that are available. These methods involve a variety of image processing techniques, for example, intensity thresholds (for details see below), edge detection (Preminger et al. 2001), watershed (Zharkov et al. 2005), morphological operations (Zharkov et al. 2005; Curto et al. 2008; Watson et al. 2009; Zhao et al. 2016), region growing (Colak & Qahwaji 2008), and level-set (Goel & Mathew 2014; Yang et al. 2018). Besides that, the statistical Bayesian method (Turmon et al. 2002) and the fuzzy-sets method (Fonte & Fernandes 2009) have been proposed.

Among the above methods, the intensity threshold method is essential because almost all these methods require thresholds to segment sunspots from the background. Usually, in a first step the outer boundaries of sunspots are detected using an intensity threshold after removing solar limb-darkening, and then the umbra-penumbra boundaries are detected using a second threshold. The suitable thresholds for segmenting the umbra and penumbra are generally derived with the aid of several image processing techniques, such as intensity profiles, intensity histograms, or morphological operations.

Early authors used a priori estimated intensity threshold to detect sunspots, for example, 15% (Chapman et al. 1994) or 8.5% (Steinegger et al. 1990) below the quiet Sun background based on the intensity distribution of several sunspots. A similar method was applied for the umbra-penumbra and penumbra-photosphere transitions at 59% and 85% of the photospheric intensity from constant intensity boundaries in Steinegger et al. (1990). Later, Beck & Chapman (1993) set a threshold from the point of maximum slope of the intensity profiles across the sunspots. Steinegger et al. (1996) and Pettauer & Brandt (1997) derived thresholds from the intersections of linear fits to the intensity cumulative histogram. Zharkov et al. (2005) and Curto et al. (2008) applied an iterative threshold method that starts from a very low intensity threshold, and then the threshold level gradually increases until the population of the detected pixels increase dramatically, indicating that the background level has been reached. Similarly, Gyori (1998) and Győri (2012) developed a set of programs named Sunspot Automatic Measurement that decompose sunspots into an ordered contour set at different intensity levels. Colak & Qahwaji (2008) set a threshold as $\mu \pm (\alpha \times \sigma)$, where μ and σ are the mean and standard deviation of the continuum images, and α was empirically set as 2.7. Later, Colak et al. (2011) set α as 2.5 after enhancing images, and Cho et al. (2015) set α as 3 using the data obtained with *Solar Dynamics Observatory* (*SDO*)/Helioseismic Magnetic Imager (HMI; Scherrer et al. 2012). Tlatov et al. (2014) and Tlatov & Pevtsov (2014) set a composite threshold by combining a relative intensity value and a gradient value. Zhao et al. (2016) set thresholds as 20% and 15% of the photospheric intensity inside and outside 0.8 solar radius of the solar disk, respectively. Yang et al. (2018) derived the threshold for umbra-penumbra boundaries using an Otsu method based on an intensity histogram.

The threshold is very critical for sunspot detections and sunspot area calculations. A decrease in threshold could miss some of the pixels that are part of the sunspots, leading to reduced sunspot areas. In contrast, an increase in threshold could increase the sunspot areas owing to some pixels being included that are not a part of the sunspots. With the evolution of solar magnetic fields, the intensities of solar images are continuously undergoing changes, especially sunspots. Therefore, self-adaptive thresholds are required. In the present study, we adopted an artificial intelligence technology called the simulated annealing genetic (SAG) method, which combined the genetic algorithm (GA) and simulated annealing (SA) algorithm to self-adaptively produce dual thresholds for detecting the umbra and penumbra of sunspots simultaneously. The paper is organized as follows. In Section 2, we provide an introduction to the GA and SA algorithm. The data and detailed steps of the SAG method are described in Section 3. In Section 4, the detection results and sunspot areas are presented and compared. In Section 5 and Section 6, the discussion and conclusion are provided, respectively.

2. Genetic Algorithm and Simulated Annealing Algorithm

2.1. Genetic Algorithm

The GA is a metaheuristic algorithm inspired by the process of natural selection that belongs to evolutionary algorithms (Charbonneau 1995). It is commonly used to generate high-quality solutions by population evolution and relies on bio-inspired operators such as selection, mutation, and crossover. The population evolution usually starts from a population with several randomly generated individuals, which is called the generation in each iteration. The fitness of every individual in the population is evaluated for each generation. The individuals are stochastically selected from the current population and the genome of each individual is modified to form a new generation by crossover and mutation. Crossover is used to produce child individuals from more than one parent individual by varying the programming of their chromosomes, and mutation is to maintain genetic diversity by altering one or more genes in a chromosome. The new generation is then evolved during the next iteration until a satisfactory fitness level has been reached for the population. The following steps are used in GA:

(1) Initialize a population with randomly generated individuals, $P_t = P_0$;
(2) Calculate the fitness of the individuals, $F(P_t)$;
(3) Select some individuals from P_t;
(4) Cross and (or) mutate the selected individuals to generate a new population P_{t+1} with a certain probability;
(5) Repeat steps (2)–(4) until the fitness of the population is stabilized or after a predefined maximum number of iterations; and
(6) Output the final solution as the individual with the best fitness.

2.2. Simulated Annealing Algorithm

The SA algorithm is a probabilistic technique for approximating the global optimum in large search spaces. The inspiration comes from annealing in metallurgy, a technique involving heating and controlled cooling of a material to increase the size of its crystals and reduce their defects (Kirkpatrick et al. 1983; Stoica et al. 2005). The key to the SA algorithm is in controlling the slow cooling that aims at a slow decrease in the probability of accepting worse solutions. It allows for escaping from the local optimal solutions and then finding the global optimal solution.

In general, the SA algorithm works as follows. The temperature T is initialized with a positive high value and the solution S_t is initialized with a random solution S_0. At each iteration, a new solution S_{new} is selected and its quality is compared by the fitness function $F(S_{\text{new}})$ with the previous S_t. The new solution will be accepted if it is better than the previous one. If not, the worse solution can also be accepted if the probability p is greater than a number that is generated randomly. The probability p is measured as

$$p = \exp\left(-\frac{F(S_{\text{new}}) - F(S_t)}{T}\right). \tag{1}$$

Then, the temperature T progressively decreases by a cooling decay factor β. At high temperatures most new solutions

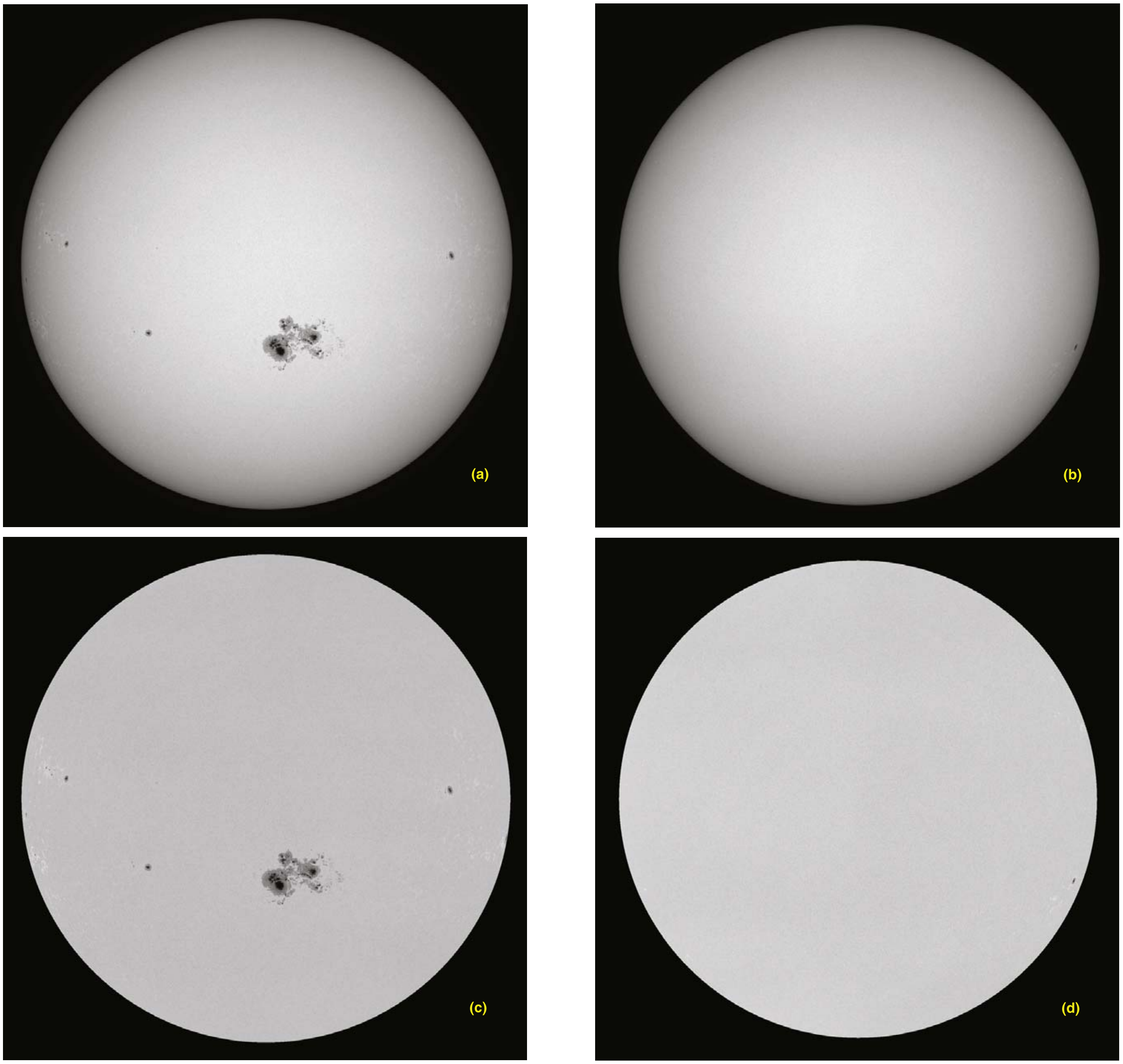

Figure 1. Two typical cases before and after removing the solar limb-darkening. (a) shows an active Sun with very large and strong sunspots in solar cycle 24, at 04:00 UT, 2014 October 24. (b) shows a very quiet Sun at 00:00 UT, 2010 July 8, where only a small sunspot is visible in the bottom right corner. (c) and (d) show the results of (a) and (b) after removing the solar limb-darkening, respectively.
(A color version of this figure is available in the online journal.)

are accepted because the accepted probability p is relatively high. When the temperature decreases, the accepted probability decreases. The role of the random number is for making worse solutions accepted more randomly and self-adaptively. If the temperature goes down slowly enough, the solutions are likely to jump out from the local optima. At convergence, when the temperature T is close to zero, the solution is frozen at the desired global optimum. The following is the common procedure for the SA algorithm:

(1) Initialize the temperature T with a positive high value, and the solution S_t with a random solution S_0;
(2) Produce a new solution S_{new};
(3) Calculate the fitness function, $F(S_t)$ and $F(S_{\mathrm{new}})$;
(4) Replace S_t with S_{new} if $F(S_{\mathrm{new}})$ is greater than $F(S_t)$ or Equation (1) is greater than a number that is generated randomly;
(5) Decrease T with a cooling decay factor β; and
(6) Repeat steps (2)–(5) until $T \leqslant 0$.

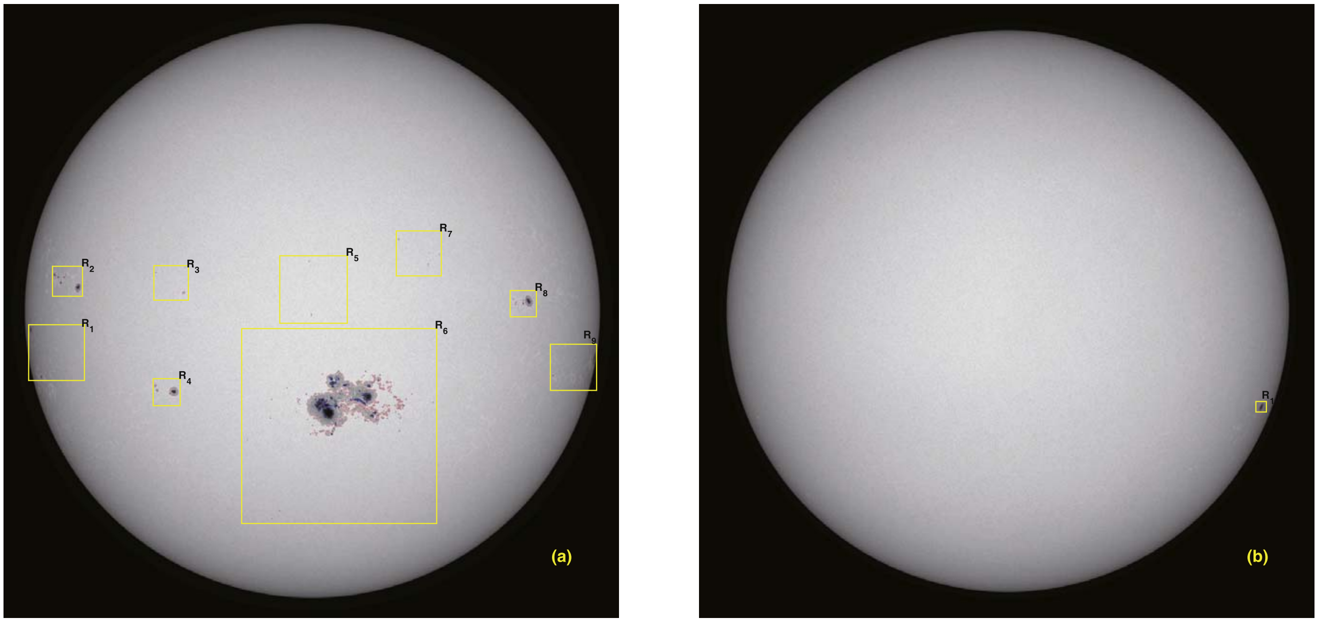

Figure 2. Detection results of Figure 1. The boundaries of the umbra-penumbra and penumbra-photosphere are contoured in blue and red, respectively. All of the sunspots are assigned to different regions.
(A color version of this figure is available in the online journal.)

2.3. SAG Method

The GA produces rich solutions by crossover and mutation; however, it usually tends to be trapped in a local optimum. Conversely, the SA algorithm is good at approximating the global optimum because of the strategy of accepting worse solutions. Therefore, we combined the virtues of both these methods. Based on the SA algorithm, the second step is modified with the new solution being produced by crossover and mutation operators. The new algorithm is referred to as the SAG method.

3. Data and Method

3.1. Data

SDO/HMI provides the highest spatial-resolution and temporal-resolution full-disk solar white-light continuum intensity images in the Fe II absorption line at 6173 Å. The spatial-resolution of these images is 1″, with a sampling of 0.″5/pixel. The temporal-resolution is 45 s (720 s for a better signal-to-noise ratio). Some corrections, such as exposure time, dark current, flat field, and cosmic-ray hits are performed in the level-1 data. Approximately 13,800 continuum intensity images with a temporal-resolution of 45 s from 2010 May to 2016 December, with a four-hour cadence were selected.

To verify the sunspot area, umbral area, and penumbral area obtained by the SAG method, we compared our results (http://61.166.157.71) with the HMI Debrecen Data (HMIDD; Baranyi et al. 2016; Győri et al. 2017), which are available at http://fenyi.solarobs.csfk.mta.hu/en/databases/SDO/. The HMIDD catalogs are detailed databases containing umbral area and penumbral area data with a temporal resolution of 1–1.5 hr from 2010 to 2014 and a temporal resolution of one day from 2015 to 2016. It provides both projected area and projection-corrected area in millionths of the solar hemisphere. In the present study, the projected area data in millionths of the solar hemisphere were used.

3.2. Preprocessing

Because of the center-to-limb variation, the image contrast deceases toward the solar limb. We removed the limb-darkening by referring to previous studies (Denker et al. 1999; Zharkova et al. 2003; Zharkov et al. 2005). The main steps were as following:

(1) Center a full-disk image;
(2) Obtain an average radial profile by computing the median value at each radial position in concentric rings (or in polar coordinates) and smoothing it to flatten the profile;
(3) Obtain a "Quiet Sun" background image by replacing each row with the average radial profile; and
(4) Obtain a flat solar image by subtracting the background image from the full-disk image.

Figure 1 shows two typical cases before and after removing the solar limb-darkening. Figure 1(a) shows an active Sun with very large and strong sunspots in solar cycle 24, at 04:00 UT, 2014 October 24. Figure 1(b) shows a very quiet Sun at 00:00 UT, 2010 July 8, in which only a small sunspot is located in the bottom right corner. Figures 1(c) and (d) show the results of (a) and (b) after removing the solar limb-darkening, respectively.

3.3. Sunspot Detection

To encode the thresholds into binary, the images after removing the solar limb-darkening were normalized in the range from 0 to $2^n - 1$, where n is suggested as an integer rounded down of log base 2 of the maximum intensity value. Then, the normalized images were smoothed by an average filtering with a 5×5 pixels^2 structuring element. Next, the dual thresholds for segmenting the umbra and penumbra were derived simultaneously with the SAG method applying the following steps.

First, several parameters were initialized. The initial temperature T, cooling decay factor β, and population size were set to $1e^5$, 0.995, and 20, respectively. The population size means that the population includes 20 individuals, e.g., from S_1 to S_{20}, and each individual includes two chromosomes, e.g., S_1 includes C_{s11}, C_{s12}, S_2 includes C_{s21}, C_{s22}. Each chromosome represents a threshold, which is encoded as n bit string separately from decimal to binary. Each bit can be regarded as a gene.

Second, the fitness values of all individuals were evaluated. The fitness function was designed following the best histogram entropy method (also known as the KSW entropy method; Kapur et al. 1980). According to the concept of Shannon entropy, suppose the intensity range of an image is from 0 to $L - 1$, its entropy H is defined as

$$H = -\sum_{i=0}^{L-1} p_i \ln p_i, \tag{2}$$

where, p_i is the appearing probability of the pixels whose values are i, expressed as $p_i = n_i/N$, where n_i is the number of pixels whose values are i, and N is the total number of image pixels. Since the continuum images were to be divided into three groups, i.e., umbra, penumbra, and photospheric granulation, dual thresholds were needed. Suppose the best dual thresholds are K_1 and K_2, then the pixels whose values are in the range of 0–K_1 compose the umbra, the pixels whose values are in the range of $K_1 + 1$ to K_2 compose the penumbra, and the pixels whose values are in the range of $K_2 + 1$ to $L - 1$ compose the granulation. The best dual thresholds maximize the entropy sum of these three groups.

The cumulative probability of umbral regions whose value was from 0 to K_1 was expressed as

$$P_0(K_1, K_2) = \sum_{i=0}^{K_1} p_i, \tag{3}$$

of penumbral regions whose value was from $K_1 + 1$ to K_2 was expressed as

$$P_1(K_1, K_2) = \sum_{i=K_1+1}^{K_2} p_i, \tag{4}$$

and of the granulation or quiet Sun whose value was from $K_2 + 1$ to $L - 1$ was expressed as

$$P_2(K_1, K_2) = \sum_{i=K_2+1}^{L-1} p_i. \tag{5}$$

The entropy of the umbral, penumbral, and granular regions were

$$H_0(K_1, K_2) = -\sum_{i=0}^{K_1} \frac{p_i}{P_0(K_1, K_2)} \ln \frac{p_i}{P_0(K_1, K_2)}, \tag{6}$$

$$H_1(K_1, K_2) = -\sum_{i=K_1+1}^{K_2} \frac{p_i}{P_1(K_1, K_2)} \ln \frac{p_i}{P_1(K_1, K_2)}, \tag{7}$$

$$H_2(K_1, K_2) = -\sum_{i=K_2+1}^{L-1} \frac{p_i}{P_2(K_1, K_2)} \ln \frac{p_i}{P_2(K_1, K_2)}, \tag{8}$$

respectively. The entropy sum is then

$$H(K_1, K_2) = H_0(K_1, K_2) + H_1(K_1, K_2) + H_2(K_1, K_2). \tag{9}$$

Then, two individuals, S_i(composed of two chromosomes C_{si1} and C_{si2}) and S_j(composed of C_{sj1} and C_{sj2}) were selected randomly to generate the next generation by crossover and mutation. The key to the crossover is swapping some genes between the parent chromosomes to generate a new generation with a certain probability. The aim of the probability was to avoid chromosomes with good fitness being crossed. We used an adaptive probability, $P_{\rm cm}$, which is defined as

$$P_{\rm cm} = \frac{\max(\{H(C_{sk1}, C_{sk2})|1 \leqslant k \leqslant 20\}) - \max(H(C_{si1}, C_{si2}), H(C_{sj1}, C_{sj2}))}{\max(\{H(C_{sk1}, C_{sk2})|1 \leqslant k \leqslant 20\}) - \mathrm{avg}(\{H(C_{sk1}, C_{sk2})|1 \leqslant k \leqslant 20\})}, \tag{10}$$

where, the functions max() and avg() measure the maximum value and average value of the parameters, respectively. The $P_{\rm cm}$ value will be relatively small if either of the chromosomes has good fitness. Otherwise, the $P_{\rm cm}$ value will be relatively large if both parents have worse fitness values. If $P_{\rm cm}$ is greater than a random number, the chromosomes, C_{si1} and C_{sj1} are

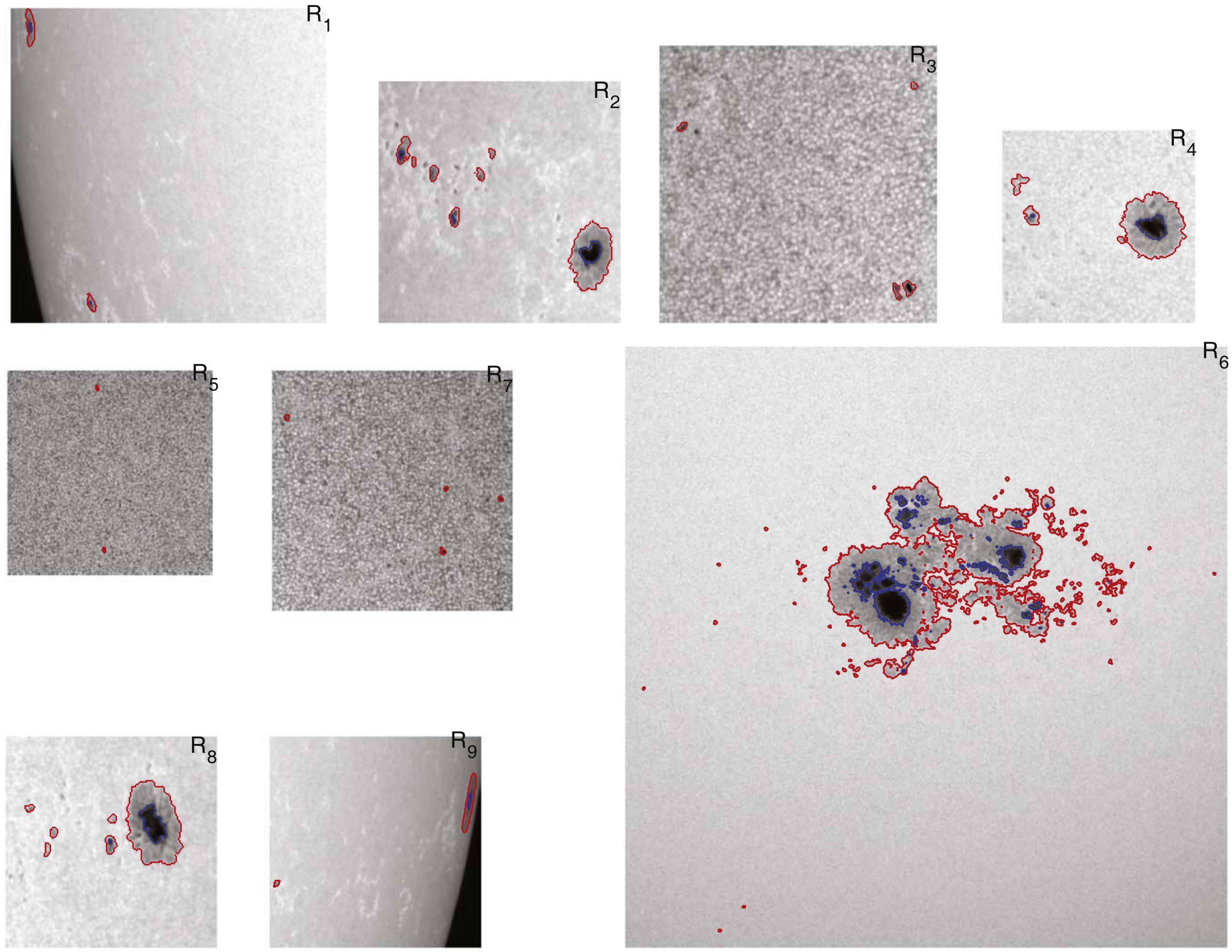

Figure 3. Nine regions labeled from R_1 to R_9, which include all sunspots of Figure 2(a). The boundaries of the umbra-penumbra and penumbra-photosphere are contoured in blue and red, respectively.
(A color version of this figure is available in the online journal.)

crossed by swapping the segments from r_1 to the end, and C_{si2} and C_{sj2} are crossed by swapping the segments from r_2 to the end, respectively. Here, r_1 and r_2 were randomly generated whose ranges were from 0 to n.

After crossover, the two new chromosomes were then mutated. Whether the genes of chromosomes were mutated or not depends on the probability P_{cm}. Similar to the crossover, the chromosome with good fitness has a small chance of mutating its genes. For each gene g_i, a new random number r_i, was generated. If the P_{cm} value is greater than r_i, the gene g_i will be mutated from 0 to 1, or from 1 to 0, otherwise it will be copied. This process yields the new generations.

The fitness values of the two new generations were re-evaluated. The old individual was replaced with the new one if the fitness of the new individual was better than that of the old one. Otherwise, a probability p was calculated by the Equation (1). If the p value was greater than a random number, the old individual was also replaced with the new individual even if the fitness of the new one was worse than that of the old one.

Until now, a new population was evolved entirely. The temperature T decreased with the cooling decay factor β. The above steps were repeated until the T value decreased to 0 or the average fitness value of all individuals was very close to the maximum fitness value, which means all individuals in the population had evolved very well.

As a result, the individual composed of two chromosomes with the best fitness was selected as the best dual thresholds. The small threshold was used to separate the umbra-penumbra boundaries and the large one was to separate the penumbra-photosphere boundaries separately.

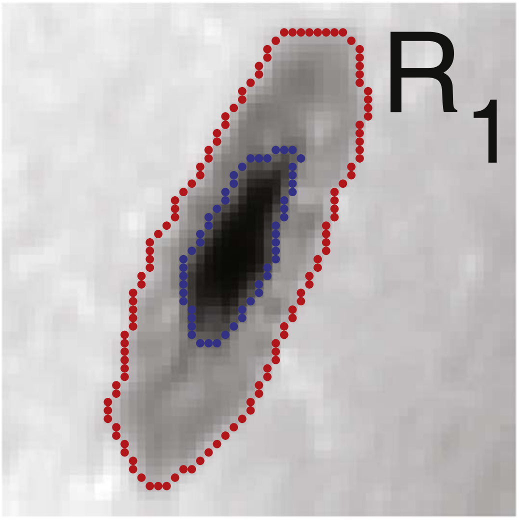

Figure 4. Sunspot region of Figure 3(b). The Sun was very quiet; only one small sunspot appeared at the solar limb.
(A color version of this figure is available in the online journal.)

Finally, candidates whose areas were smaller than 2 millionths of the solar hemisphere (μHem) were removed following with Watson et al. (2009) or Cho et al. (2015).

4. Results

4.1. Sunspots

Figure 2 shows the detection results of Figure 1. The boundaries of the umbra-penumbra and penumbra-photosphere are contoured in blue and red, respectively. All the sunspots (including some pores) were detected perfectly, even those sunspots located near the extreme solar limb.

To check the performance of sunspot detection in greater detail, all the sunspots in Figure 2(a) were assigned to nine regions labeled from R_1 to R_9, and zoomed in Figure 3. It can be seen that the contours of the umbra and penumbra are delineated finely. In particular, the sunspots located at the extreme solar limb in R_9 present clear contours. Some slight dark candidates that could be seen by the naked eyes have not been marked because of their very small size (less than 2 μHem). The minimum size of sunspots will be discussed in Section 5.

Figure 4 shows the sunspot region of Figure 3(b). The Sun was very quiet, only one small sunspot appeared at the solar limb. Nonetheless, both the umbra and penumbra are very well separated.

4.2. Sunspot Areas

The sunspot area is an important indicator of the solar activity level (Carbonell & Ballester 1992), which is also associated with the solar cycle and is significant in space weather monitoring. Table 1 lists the total sunspot areas, umbral areas, and penumbral areas in Figures 1(a) and (b) obtained by the SAG method and HMIDD separately. The sunspot area in Figure 1(a) obtained from the SAG method is $\sim$1800 μHem larger than that obtained from the HMIDD. The difference is very large, however, we feel confident in the detection results by the SAG method from Figures 2(a), 3(a), and 7. The detail will be discussed in Section 5. The areas of Figure 1(b) between the SAG method and HMIDD were very close, within a few percent.

Figure 5 plots the daily average sunspot areas extracted by the SAG method and HMIDD during the period from 2010 May to 2016 December. The horizontal axis represents the date the images was taken and the vertical axis represents the total sunspot areas, umbral areas, and penumbral areas in units of μHem. Note the different coordinate ranges of vertical axes in the seven panels. The sunspot areas during the second half of 2010 ranged from 0 to 1467 μHem. The Sun was in its minimum and there were even some spotless days. Over the next three years, the sunspot areas ranged from $\sim$10 to $\sim$5000 μHem. In 2014, one of the biggest sunspots in history evolved, with the sunspot areas varying from 96 to 10,344 μHem during this year, owing to NOAA 12192. The next two years' solar activity decreased again, the sunspot areas being ranging from 22 to 3902 μHem and from 0 to 2263 μHem.

The inner structure of sunspots is important to know the model of sunspots (Osherovich & Garcia 1989; Li et al. 2018). The ratio of umbral area to total sunspot area, $r_u = S_u/S_s$, is a good parameter. The mean ratios per year from 2010 to 2016 were 0.233, 0.174, 0.168, 0.164, 0.157, 0.174, and 0.221, respectively. The ratio of 2014 was slightly smaller than the other years and the ratios of 2010 and 2016 were slightly larger than the other years. The ratio decreased with an increase in solar activity. During the solar minimum, almost no big sunspot group types such as E or F appeared, with most of the sunspots being small or unipolar. This is the main reason for the differences seen. Previous studies have reported the ratio r_u, for example, the ratio was measured as $\sim$0.17 (Tandberg-Hanssen 1956; Gokhale & Zwaan 1972; Antalova 1991); later, the ratio was reported as $\sim$0.2 (Beck & Chapman 1993; Martinez Pillet & Vazquez 1993; Tlatov et al. 2014). Several authors have attempted to determine the relationship between the ratio and other parameters such as sunspot size and solar cycle. Jensen et al. (1955) found 0.19 around the maximum of the sunspot cycle and 0.16 around the minimum. The ratio increased slightly with increasing sunspot size (Steinegger et al. 1990; Beck & Chapman 1993), e.g., from 0.19 to 0.24 for small and large sunspots, respectively. The difference between the results probably relates to the different techniques applied to measure the umbral and penumbral areas and the quality of the data (Solanki 2003). The ratios we obtained were in the same range as those from previous studies.

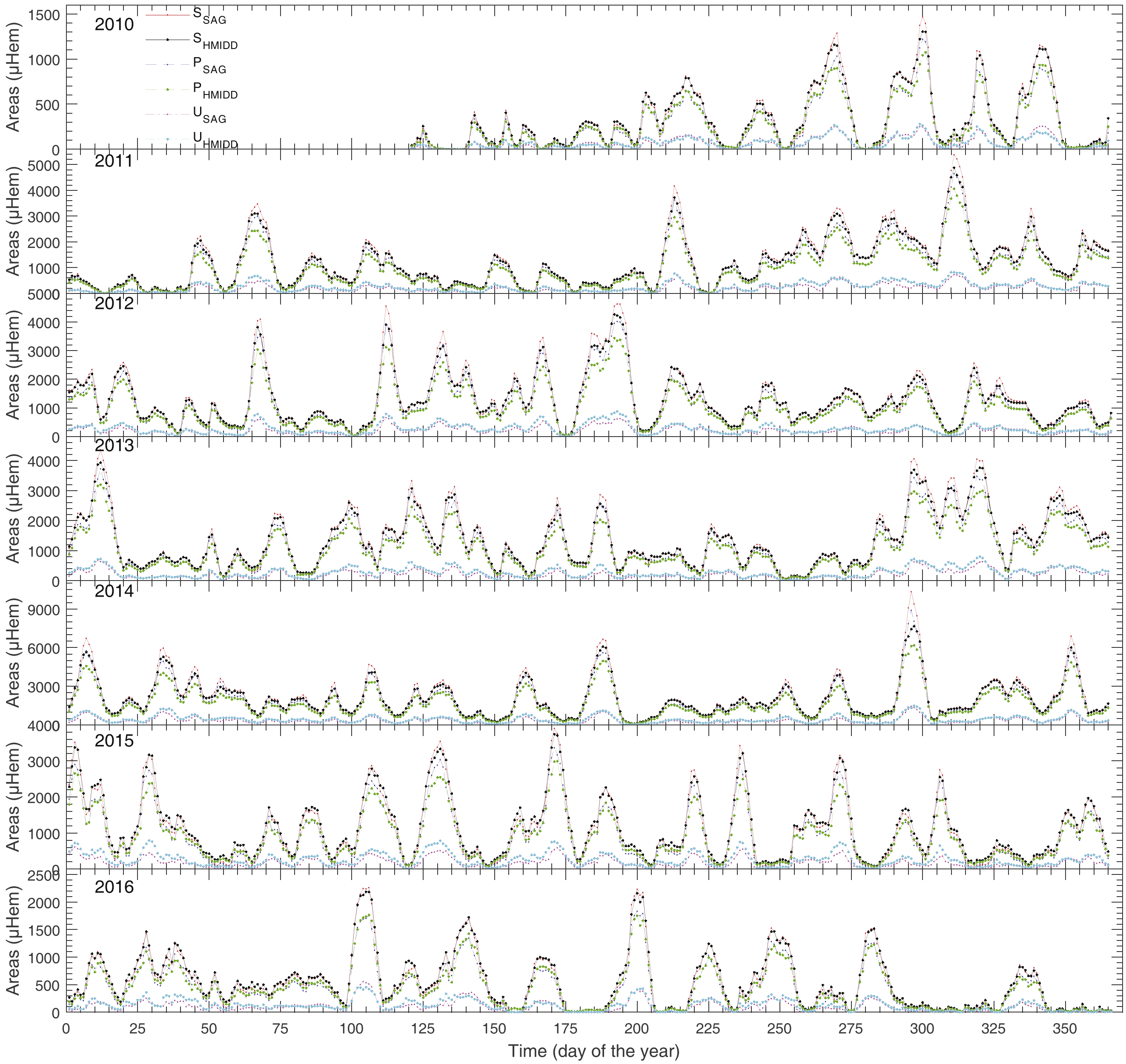

Figure 5. Daily average sunspot areas extracted by the SAG method and HMIDD during the period from 2010 May to 2016 December. The horizontal axis represents the date the images was taken, and the vertical axis represents the areas in units of μHem.
(A color version of this figure is available in the online journal.)

Table 1
The Total Sunspot Areas of Figure 1

	Figure 1(a)			Figure 1(b)		
	Sunspot Areas (μHem)	Umbral Areas (μHem)	Penumbral Areas (μHem)	Sunspot Areas (μHem)	Umbral Areas (μHem)	Penumbral Areas (μHem)
SAG	10052	1415	8637	79	15	64
HMIDD	8223	1559	6664	81	13	68

Table 2
The Correlation Coefficients of the Sunspot, Umbral, and Penumbral Areas between SAG and HMIDD

Year	Correlation Coefficient of Sunspot Areas	Correlation Coefficient of Umbral Areas	Correlation Coefficient of Penumbral Areas
2010	0.996	0.992	0.958
2011	0.998	0.996	0.957
2012	0.996	0.994	0.951
2013	0.997	0.994	0.939
2014	0.992	0.990	0.962
2015	0.994	0.991	0.955
2016	0.996	0.992	0.924
Avg.	0.996	0.993	0.949

The total sunspot areas, umbral areas, and penumbral areas between the SAG method and HMIDD were in very good agreement. Table 2 lists the correlation coefficients of these areas between the SAG method and HMIDD. All correlation coefficient values were very high, e.g., the values of the sunspot areas and penumbral areas were all as high as 0.99, and the values of umbral areas were all above 0.92, with the average value being 0.949.

The total sunspot areas between SAG and HMIDD were very consistent for the majority of the time; however, the areas of SAG were slightly larger than those of HMIDD on most peak days. We calculated the area differences of each day between SAG and HMIDD. The maximum differences of each year were less than several hundred μHem, except $\sim$3000 μHem on 2014 October 23. Figures 6(a) and (b) show the image at 04:00 UT, 2014 October 23 and the detection result, respectively, which had the largest total sunspot area difference and the largest penumbral area difference among all the images. The total sunspot area and penumbral area of SAG were 3459 and 3402 μHem larger than those of HMIDD, respectively. We checked the detection results and the data of the sunspot areas. The total sunspot areas obtained by the SAG method at 00:00, 04:00, and 08:00 UT, 2014 October 23 were 10,211, 10,282, and 10,344 μHem, respectively, and those recorded for HMIDD were 7519, 6823, and 7548 μHem, respectively. The successive images show the continuous evolution of sunspots with the data of HMIDD appearing to be discontinuous. Figures 6(c) and (d) show the image at 20:00 UT, 2014 February 6 and the detection result, respectively, which had the largest umbral area difference among all the images. The umbral area of SAG was 491 μHem smaller than that of HMIDD.

5. Discussion

5.1. Parameters

The SAG method aims to derive suitable dual thresholds automatically. A few parameters still need to be set, although most parameters are generated randomly. The initial temperature and cooling decay factor play crucial roles in controlling the probability of accepting worse solutions. By cooling the temperature slowly enough, the global optimum can be found, however, the computing time will increase. The suitable parameter values can be found experimentally, we tried with temperatures in the range of $1e^3$–$1e^6$, and the cooling decay factor in the range of 0.9–0.9999. The result was that the thresholds were quite stable if the temperature and cooling decay factor were no less than $1e^5$ and 0.995, respectively. The average computing time of one image is less than two seconds (using a normal PC). Additionally, the computing time increases only slightly even if both parameter values increase, because the evolution process will be terminated when the entire population has already been evolved perfectly. Conversely, the evolved thresholds become unstable if initial numbers less than $1e^3$ and 0.99 are used. Therefore, the initial temperature and cooling decay factor are suggested to be in the range of $1e^4$–$1e^5$ and 0.99–0.999, respectively. Besides that, the population size also needs to be set. This will affect the evolving time because population evolution can be terminated when all individuals in the population are well evolved. The larger the population size is, the more likely that the global optimum will be found, and the more evolving time costs. The experimental results showed that it was suitable for setting in the range of 20–40.

Different rules for distinguishing the smallest sunspots have been used by previous authors. Using SOHO/MDI data with a pixel resolution of 2″, Zharkov et al. (2005) and Goel & Mathew (2014) did not exclude any small sunspots; Curto et al. (2008) only selected sunspots larger than 7 × 7 MDI pixels corresponding to 35 μHem; and Watson et al. (2009) removed the candidates whose areas were less than 30 MDI pixels corresponding to 20 μHem. Using the China Huairou Solar Observing Station data with a pixel resolution of 2″, Zhao et al. (2016) took into account sunspots exceeding 2° in diameter corresponding to $\sim$30 μHem. Recently, Cho et al. (2015) separated the candidates into three groups as follows: pores with an area smaller than 20 μHem, transitional sunspots with an area from 20 to 100 μHem, and mature sunspots with an area exceeding 100 μHem. They excluded the candidates with an area smaller than 2 μHem using *SDO*/HMI data with a pixel resolution of 0.″5.

We performed area calculations by removing sunspots smaller than 0, 1, 2, 3, 4, 5, 10, 20, and 30 μHem. The total sunspot areas in Figure 1(a) correspond to 10,312, 10,198, 10,052, 9964, 9869, 9818, 9700, 9572, and 9447 μHem, respectively. They are all larger than the value of HMIDD (8223 μHem). Figure 7 shows the results of Figure 1(a) after excluding the candidates whose areas were less than 0, 2, 5, and 20 μHem. The solar image is densely covered with lots of sunspots, pores, or very tiny dark spots in Figure 7(a), whereas only large and clear sunspots are left in Figure 7(d). Even so, the total sunspot area in Figure 7(d) is larger than that of

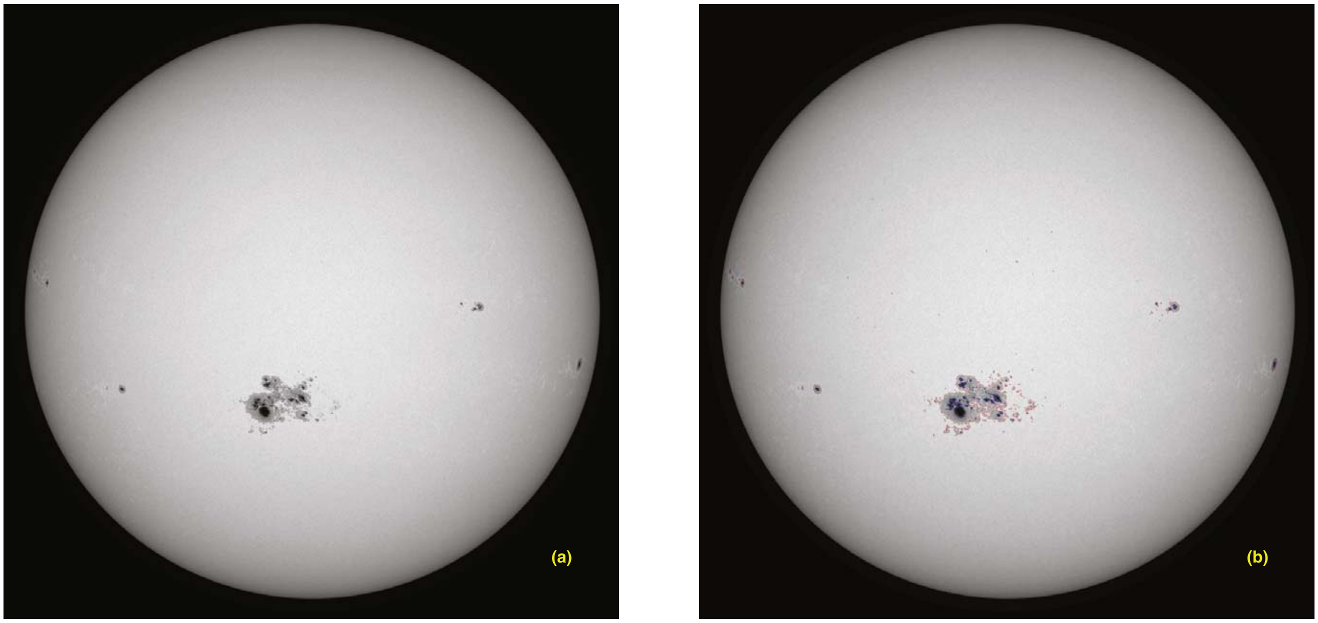

Figure 6. Detection results of images with maximum area difference. (a) and (b) show the image at 04:00 UT, 2014 October 23 and the detection result, respectively, which had the largest total sunspot area difference and the largest penumbral area difference among all the images. (c) and (d) show the image at 20:00 UT, 2014 February 6 and the detection result, respectively, which had the largest umbral area difference among all the images.
(A color version of this figure is available in the online journal.)

HMIDD beyond ~1200 μHem. Therefore, to avoid some noise and ephemeral magnetic signals, we excluded those candidates whose sizes were less than 2 μHem following Cho et al. (2015).

5.2. *Improving the SAG Method for Image Series*

The computing time of the SAG method can be improved dramatically when a time series of full-disk solar images is processed. The successive images correlate well because of the continuous evolution of magnetic fields, especially the high cadence of image series. This will lead to similar thresholds of successive images. Figure 8(a) plots the dual thresholds of all images in 2014 using the SAG method with blue solid line and red dashed–dotted line (the images are normalized from 0 to $2^{15} - 1$). Remember the sunspot areas in 2014 varied a lot. Even so, the threshold for segmenting the penumbra-photosphere varied very slowly, with the difference between two successive images being less than 2% of image gray range and the variation over the entire year being limited to a range of 3%. The threshold for segmenting the umbra-penumbra varied greatly, with the largest difference between two successive images being ~10% and the variation over the entire year was up to ~25%. Therefore, the SAG method can be revised to make the thresholds evolve within a limited range. In the procedure of the SAG method, the initial population is generated randomly and the next generation is generated by crossover and mutation. Because the successive images of a time series are correlated, the initial population from the second image can be generated from the dual threshold values of the prior image and the next population can be generated with only a small adjustment. The dual thresholds of the first image, tl_1 and th_1, were still produced using the SAG method. From the second image, we generated the initial population of each image limited in the range of $\pm tl_1 \times 1.1$ and $\pm th_1 \times 1.02$ randomly. Besides that, we abandoned crossover in the step of generating the next population, and only conserved the mutation to fine adjust the threshold values. The tl_1 value was mutated by changing its random bits in the range from the 1st bit to the 12th bit on the right, where the threshold value can be varied in the range of 13% ($2^{12} - 1 = 4095$). The th_1 value is mutated by changing its random bits in the range from the 1st bit to the 10th bit on the right, where the threshold value can be varied in the range of 3%. The threshold value will increase by 1 if the 1st bit on the right is changed from 0 to 1, and the value will decrease by 1 if it is changed from 1 to 0. The value will increase by 512 (2^9) if the 10th bit on the right is changed from 0 to 1, and decrease by 512 if it is changed from 1 to 0, and so on.

Figure 8(b) plots the residuals of dual thresholds between the improved SAG method and the SAG method using a blue solid line and red dashed–dotted line, respectively. Ideally the residuals ought to be the same. However, these residuals were very small because of the approximate optimal solution derived by either the GA or SA algorithm. The residuals were so small that the detection results were almost the same. The average

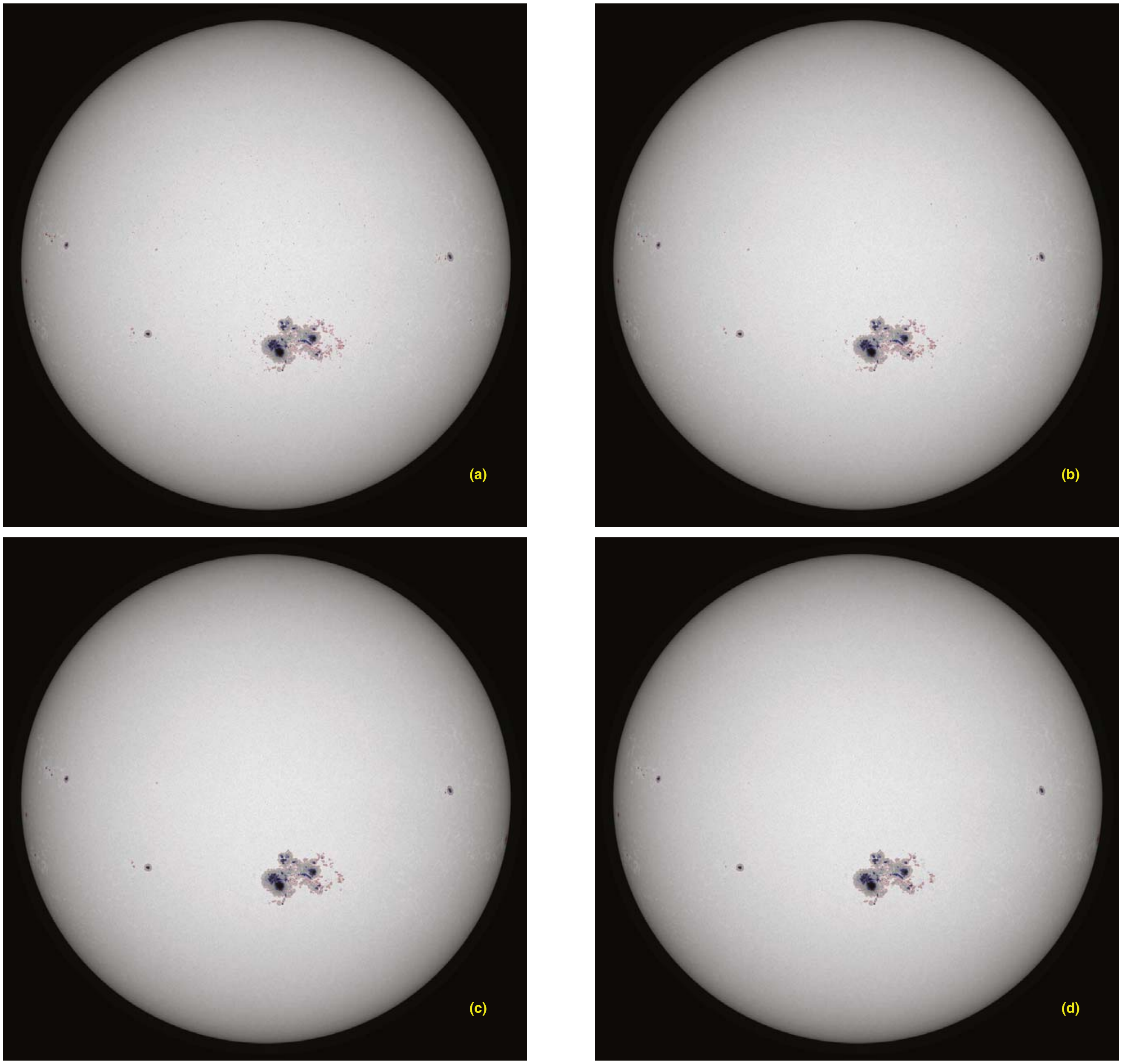

Figure 7. Detection results of Figure 1(a) after excluding the candidates whose sizes were less than 0 μHem (a), 2 μHem (b), 5 μHem (c), and 20 μHem (d). (A color version of this figure is available in the online journal.)

computing time of an image by the improved SAG method was around one fifth that of the SAG method. The mutated bit number also depended strongly on the image cadence, i.e., the shorter the time interval between the images, the less mutation was required and the less computing time was needed.

5.3. Comparison with the Top-hat Transform Method

Image normalization in the preprocessing step is often necessary if the threshold method is applied because of solar limb-darkening (Zharkova et al. 2003). There are two typical methods to remove solar limb-darkening. One is our method using a smoothed average radial profile similar to that used by Zharkova et al. (2003) or Zharkov et al. (2005). Another is using a Top-hat transform method that is also suitable for enhancing structures that have a detectable contrast against their local neighborhood (Curto et al. 2008; Watson et al. 2009; Zhao et al. 2016). The Top-hat transform method is applied by subtracting the closed image from the original image. The closing operator can remove features (sunspots or granulation

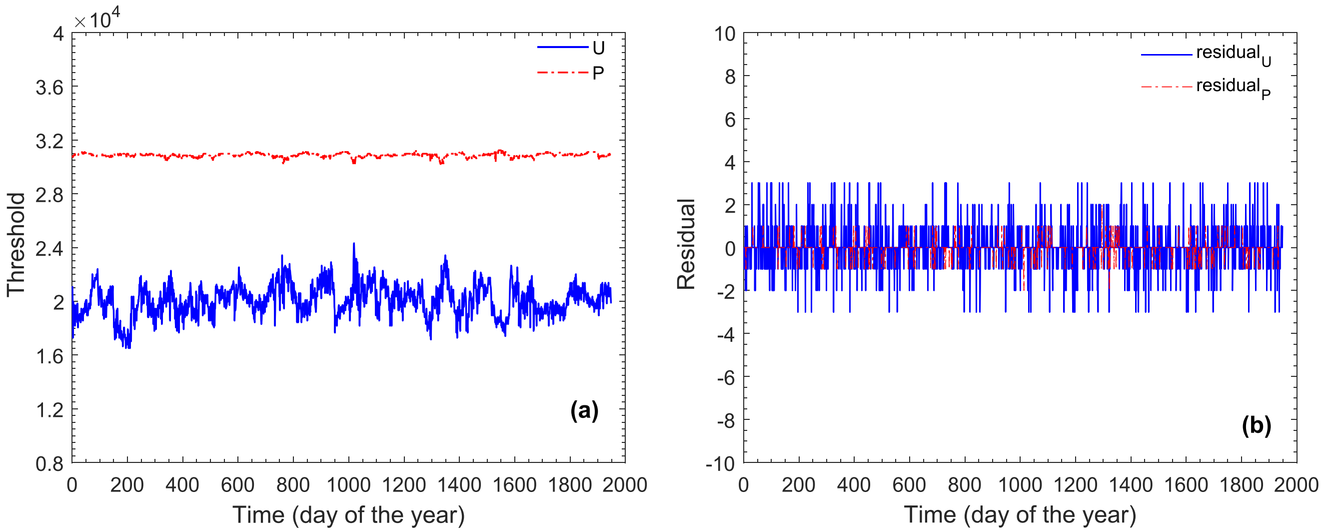

Figure 8. (a) Dual thresholds of all images in 2014 using the SAG method with a blue solid line for the umbra-penumbra transition and a red dashed–dotted line for the penumbra-granulation transition (the images are normalized from 0 to $2^{15} - 1$). (b) The residuals of the dual thresholds between the improved SAG method and SAG method.

(A color version of this figure is available in the online journal.)

within sunspot groups) smaller than the structuring element, for details please refer to Curto et al. (2008).

We compared these two methods. Figures 9(a) and (b) show the detection results of Figure 1(a) by the method we adopted in the present study and the Top-hat transform method, respectively. The total sunspot areas were 10,052 and 9177 μHem, respectively, i.e., a difference of $\sim$900 μHem. The largest sunspot regions in Figures 9(a) and (b) are zoomed in Figures 9(c) and (d), respectively. The contours of sunspots in Figure 9(c) are finer than those in Figure 9(d). The contours of sunspots in Figure 9(d) are over-smoothed. The closing operator is defined as a dilation followed by an erosion, whose effect is mainly deleting small dark structures. However, it inevitably results in smoothed borders of the remaining objects and filled grooves, which are shown in Figure 9(d). Therefore, it is necessary to be careful using mathematical morphology operators if the fine structures of objects need to be conserved.

6. Conclusion

The threshold method is essential for sunspot detection and sunspot area calculations because sunspots are visibly darker than the photosphere in continuum images. In the present study, we proposed an artificial intelligence technology called the SAG method, which combines the GA and SA algorithm to self-adaptively derive dual thresholds for detecting the umbra and penumbra of sunspots simultaneously. Approximately 13,800 full-disk continuum intensity images recorded with *SDO*/HMI were used, with a cadence of four hours from 2010 May to 2016 December. During the preprocessing step, we removed the solar limb-darkening by an average radial profile with the median value at each radial position. Then, the best dual thresholds were derived by the SAG method, which were obtained by finding the maximum entropy sum of umbra, penumbra, and background. In the procedure of SAG, new solutions are generated by population evolution relying on bio-inspired operators like selection, mutation, and crossover. The old solutions are replaced with high-quality solutions with good fitness and the worse solutions are also accepted to escape from local optimal solutions with an accepting probability. The accepting probability is controlled by slow cooling that aims at slow decreases in the probability of accepting worse solutions. After excluding some small candidates, the umbra and penumbra of sunspots were finally detected simultaneously. The detection results show that the dual thresholds derived from the SAG method are suitable for accurately and efficiently segmenting the umbra and penumbra from the photosphere, regardless of solar activity. The boundaries of the umbra and penumbra were very finely delineated, even sunspots at the extreme solar limb are resolved. Additionally, the sunspot areas were measured, which represent a more meaningful physical parameter, being more closely related to the magnetic flux. The mean ratios of umbra to sunspot areas per year were in the range of 0.159–0.233. The ratios decreased with an increase in solar activity. During the solar minimum, most of the sunspots were small or unipolar, almost no big sunspot group types such as E or F appeared. This is the main reason for the difference in the ratio value and implies that the ratio was related to the solar activity level. Comparing the total sunspot areas, umbral areas, and penumbral areas obtained by the SAG method with the reported areas of HMIDD showed very good correlation. The

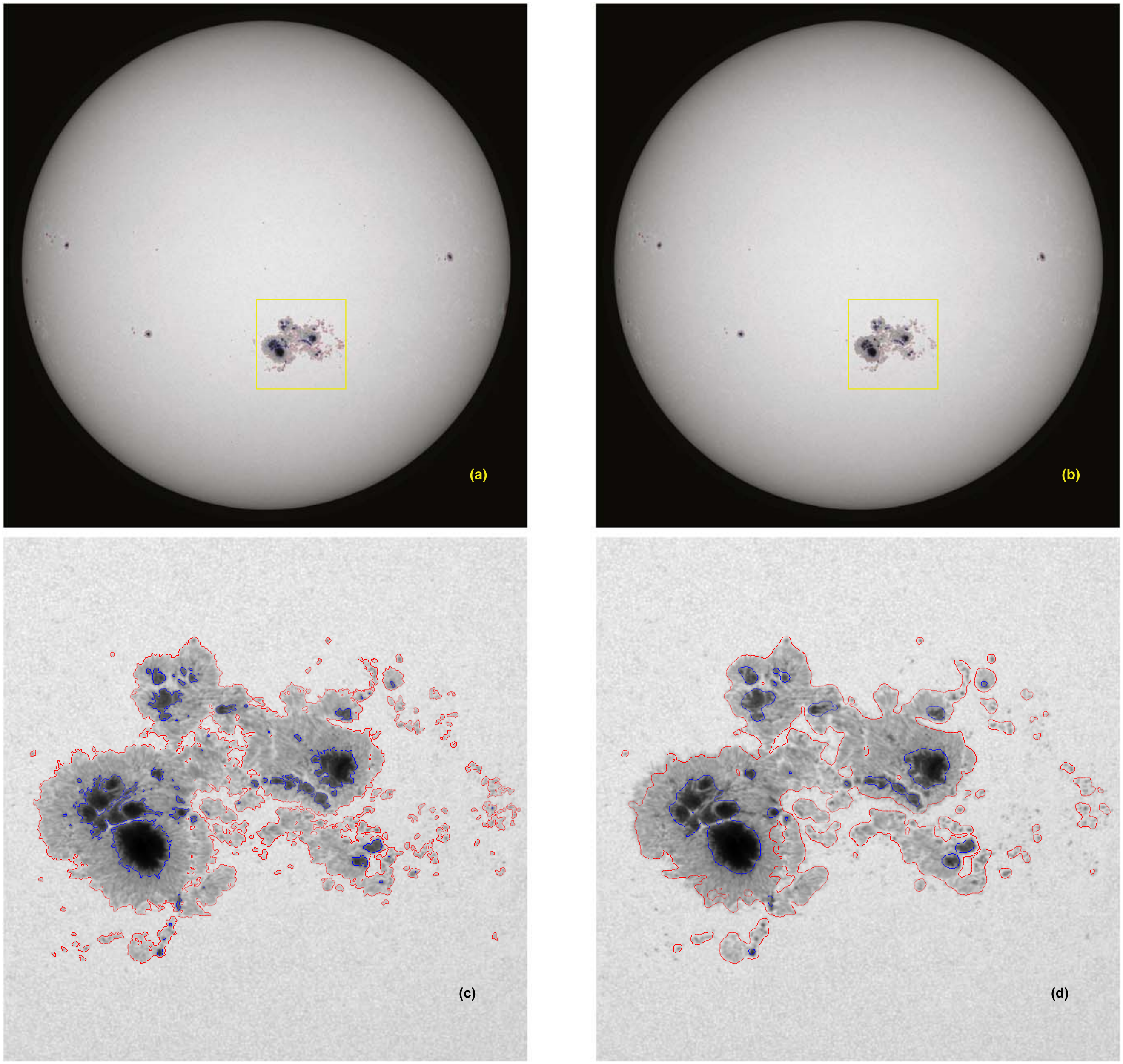

Figure 9. Detection results of Figure 1(a) by the method we adopted in the present study (a) and Top-hat operator (b), respectively. The largest sunspot regions in (a) and (b) are zoomed in (c) and (d), respectively.
(A color version of this figure is available in the online journal.)

correlation coefficients were up to 0.99, 0.99, and 0.95, respectively. In summary, the SAG method is good at detecting fine structures of the umbra and penumbra. It can provide solar parameters immediately that are very valuable for real-time space weather predictions, such as positions, intensities, and areas.

The authors thank the referee very much for the careful reading and constructive comments. The authors are grateful for the support received from the National Natural Science Foundation of China (Nos. 11763004, 11573012, 11303011, and 11463003) and the Open Research Program of the Key Laboratory of Solar Activity of the Chinese Academy of Sciences (Nos. KLSA201815). The full-disk images used in the study were kindly provided by NASA/*SDO* and the HMI science team. Thanks to the staff of Debrecen Observatory for maintaining and organizing the HMIDD catalog.

References

Antalova, A. 1991, BAICz, 42, 316
Baranyi, T., Győri, L., & Ludmány, A. 2016, SoPh, 291, 3081
Beck, J. G., & Chapman, G. A. 1993, SoPh, 146, 49
Carbonell, M., & Ballester, J. L. 1992, A&A, 255, 350
Chapman, G. A., Cookson, A. M., & Dobias, J. J. 1994, ApJ, 432, 403
Charbonneau, P. 1995, ApJS, 101, 309
Cho, I.-H., Cho, K.-S., Bong, S.-C., et al. 2015, ApJ, 811, 49
Colak, T., & Qahwaji, R. 2008, SoPh, 248, 277
Colak, T., Qahwaji, R., Ipson, S., & Ugail, H. 2011, AdSpR, 47, 2092
Curto, J. J., Blanca, M., & Martínez, E. 2008, SoPh, 250, 411
Denker, C., Johannesson, A., Marquette, W., et al. 1999, SoPh, 184, 87
Fonte, C. C., & Fernandes, J. 2009, SoPh, 260, 21
Goel, S., & Mathew, S. K. 2014, SoPh, 289, 1413
Gokhale, M. H., & Zwaan, C. 1972, SoPh, 26, 52
Gyori, L. 1998, SoPh, 180, 109
Győri, L. 2012, SoPh, 280, 365
Győri, L., Ludmány, A., & Baranyi, T. 2017, MNRAS, 465, 1259
Jensen, E., Nordø, J., & Ringnes, T. S. 1955, ApNr, 5, 167
Kapur, J. N., Sahoo, P. K., & Wong, A. C. 1980, SigPr, 2, 3
Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. 1983, Sci, 220, 671
Li, Q., Yan, X., Wang, J., et al. 2018, ApJ, 857, 21
Martinez Pillet, V., & Vazquez, M. 1993, A&A, 270, 494
Osherovich, V. A., & Garcia, H. A. 1989, ApJ, 336, 468
Pettauer, T., & Brandt, P. N. 1997, SoPh, 175, 197
Preminger, D. G., Walton, S. R., & Chapman, G. A. 2001, SoPh, 202, 53
Scherrer, P. H., Schou, J., Bush, R. I., et al. 2012, SoPh, 275, 207
Solanki, S. K. 2003, A&ARv, 11, 153
Steinegger, M., Brandt, P. N., Pap, J., & Schmidt, W. 1990, Ap&SS, 170, 127
Steinegger, M., Vazquez, M., Bonet, J. A., & Brandt, P. N. 1996, ApJ, 461, 478
Stoica, R. S., Gregori, P., & Mateu, J. 2005, Stoch. Process. Appl., 115, 11
Tandberg-Hanssen, E. 1956, ApNr, 5, 207
Tlatov, A. G., & Pevtsov, A. A. 2014, SoPh, 289, 1143
Tlatov, A. G., Vasil'eva, V. V., Makarova, V. V., & Otkidychev, P. A. 2014, SoPh, 289, 1403
Turmon, M., Pap, J. M., & Mukhtar, S. 2002, ApJ, 568, 396
Watson, F., Fletcher, L., Dalla, S., & Marshall, S. 2009, SoPh, 260, 5
Yan, X. L., Priest, E. R., Guo, Q. L., et al. 2016, ApJ, 832, 23
Yan, X. L., & Qu, Z. Q. 2007, A&A, 468, 1083
Yan, X. L., Wang, J. C., Pan, G. M., et al. 2018a, ApJ, 856, 79
Yan, X. L., Yang, L. H., Xue, Z. K., et al. 2018b, ApJL, 853, L18
Yang, M., Tian, Y., & Rao, C. 2018, SoPh, 293, 15
Zhao, C., Lin, G., Deng, Y., & Yang, X. 2016, PASA, 33, e018
Zharkov, S., Zharkova, V. V., & Ipson, S. S. 2005, SoPh, 228, 377
Zharkova, V. V., Ipson, S. S., Zharkov, S. I., et al. 2003, SoPh, 214, 89