Chinese Sunspot Drawings and Their Digitization – (I) Parameter Archives

G.H. Lin1,2 · X.F. Wang1,2 · S. Liu1,2 · X. Yang1,2 · G.F. Zhu1,2 · Y.Y. Deng1,2 · H.S. Ji3 · T.H. Zhou4 · L.N. Sun4 · Y.L. Feng5 · Z.Z. Liu5 · J.P. Tao5 · M.X. Ben5 · J. Lin5 · M.D. Ding6,7 · Z. Li6,7 · S. Zheng8 · S.G. Zeng8 · H.L. He8 · X.Y. Zeng8 · Y. Shu8 · X.B. Sun8

Received: 16 November 2018 / Accepted: 10 May 2019 / Published online: 19 June 2019
© Springer Nature B.V. 2019

Abstract Based on the Chinese historical sunspots drawings, a data set consisting of the scanned images and all their digitized parameters from 1925 to 2015 have been constructed. In this paper, we briefly describe the developmental history of sunspots drawings in China. This paper describes the preliminary handling processes that start from the initial scanning to the extraction of parameters, and finally summarizes the general features of this data set. It is the first systematic project in the Chinese solar-physics community to digitize the historical series of sunspot drawings. Our data set fills in an almost 90-year historical gap, which originates from an area spanning 60 degrees from east to west and 50 degrees from north to south and has no continuous and detailed digital sunspot observation information. Being complementary to other sunspot observations in the world, our data set provides abundant information to the long-term solar activity cycles research.

Keywords Sunspot · Solar cycle · Sunspot drawings · Digitalization · Physical parameters · Big data

S. Liu
lius@nao.cas.cn
G.H. Lin
lgh@nao.cas.cn

1 Key Laboratory of Solar Activity, Datun Rd. 20A, Chaoyang District, Beijing, 100101, P.R. China
2 National Astronomical Observatories, CAS, Datun Rd. 20A, Chaoyang District, Beijing, 100101, P.R. China
3 Purple Mountain Observatory, CAS, 2 Beijing Xi Road, Nanjing, Jiangsu, 210008, P.R. China
4 Qingdao Observatory, Purple Mountain Observatory, CAS, Guanxiang Rd. 2, Shinan District, Qingdao, 266003, P.R. China
5 Yunnan Astronomical Observatory, CAS, 396 Yanfangwang, Guandu District, Kunming, Yunnan, 650216, P.R. China
6 School of Astronomy & Space Science, Nanjing University, 22 Hankou Road, Gulou District, Nanjing, Jiangsu, 210093, P.R. China
7 Key Laboratory for Modern Astronomy and Astrophysics, Nanjing University, Ministry of Education, Nanjing 210023, P.R. China
8 College of Science, China Three Gorges University, Yichang 443002, P.R. China
Table 1  Current major sunspot information databases in the world.

| Database                                      | Period              | Ref.                                                                 |
|-----------------------------------------------|---------------------|----------------------------------------------------------------------|
| A catalogue of sunspot observations covering  | 165 BC – AD 1684    | Wittmann and Xu (1987)                                               |
| the period 165 BC to AD 1684                  |                     |                                                                      |
| University of Extremadura (UE, Spain)         | 1610 – 2010         | http://haso.unex.es/haso/                                            |
| Greenwich Photoheliographic Results (GPR, UK) | 1872 – 1976         | http://fenyi.solarobs.csfk.mta.hu/GPR/Willis et al. (2013)           |
| Kodaikanal Solar Observatory (KK, India)      | 1906 – 1987         | Howard, Gilman, and Gilman (1984) Howard (1991) Sivaraman, Gupta, and Howard (1993) Ravindra et al. (2013) |
| Mount Wilson Observatory (MW, USA)            | 1917 – 1985         | Howard, Gilman, and Gilman (1984) Howard (1991) Sivaraman, Gupta, and Howard (1993) Ulrich et al. (2004) http://www.astro.ucla.edu/~ulrich/MW_SPADP/index.html |
| Debrecen Photoheliographic Data (DPD, Hungary) | 1974 – present      | Baranyi et al. (2001) Baranyi, Györi, and Ludmány (2015)             |

1. Introduction

The sunspot observations in the past provide the most direct data resource for studying the changes in solar activity. The conventional sunspot observations not only demonstrate the variation of sunspots, but also reveal some hints to understand other solar activity phenomena such as solar magnetic field, solar rotation, white-light flares, etc. The formation of sunspots leads to the formation of active regions, which are the strongest magnetic fields in the solar photosphere and are formed through magnetic flux emergence from under the photosphere. Due to the presence of strong magnetic field, active regions containing sunspots are the main locations where violent eruptions such as solar flares can occur (Li, 2015). The intensive solar activity could have an impact on the Earth’s magnetosphere and ionosphere, e.g. telecommunication could be seriously hindered or even suddenly interrupted for a while, which can present a serious threat to the high technology system safety of aircraft, ships, and satellites, as well as telecommunication, data transfer, and so on.

The long period temporal and spatial distribution evolution of sunspots reflects the process of magnetic dynamo from an interior of the Sun. The research of solar dynamo problem relies heavily on long-term historical data accumulated by a large number of human beings. Therefore, it is important to improve the time zone coverage of the observations and the accuracy of the extracted information from historical observational data, for understanding and predicting the long-term solar behavior, and also for studying the Sun’s influence on the solar–terrestrial space environment and our human activities (Tang, 2014, 2015).

The historical sunspot observations in the world are widely spread over different longitudes of the globe. The major sunspot information databases are given in Table 1. In addition to these sunspot information databases given in Table 1, the World Data Center-SILSO (Vanlommel et al., 2004; Clette et al., 2014, http://www.sidc.be/silso/) presents the total sunspot number (daily total sunspot number [1/1/1818 – present], monthly mean total sunspot number [1/1749 – present], yearly mean total sunspot number [1700 – present]) and the number of sunspot groups from 1610 to 2010. In this paper, our Chinese sunspot database is presented to fill the longitudinal gap in long-term observations of sunspots from 1925 up to now to a great extent.
Table 2  Brief history of Qingdao Observatory Station.

| Year Range | Description |
|------------|-------------|
| 1937–1945 | Observation continued during the Japanese occupation. After the war, part of data, the 16 cm objective lens and photographic devices were plundered. |
| 1947      | Sunspots drawing resumed with a 4-inch lens and 14.4 cm solar diameter on the projection board. |
| 1954      | Objective lens was updated to 15 cm. Focal length 2.2 m. Projected solar diameter is the national uniform standard 17.4 cm. |
| 1978–1983 | Observatory was rescinded. Observation stopped. |
| 1983–1988 | Supported by Chinese Academy of Sciences, observation was resumed using a modified 20 cm aperture guide telescope with a 3.5 m focal length and solar projection diameter 17.4 cm. |
| 1988–present | 32 cm refractive telescope equipped with sunspots fine structure camera. The penumbra fibril resolution reaches 0.7”. The penumbra of sunspots in the drawings can be seen. |

Ancient China attached great importance to astronomical records because the Emperor’s divine rights were explained to be granted by Heaven. Actually, China seems to be the first country in the world to record sunspots (traceable around 364 B.C., Leo, 2011) and has long-term text records of sunspots, however, scientific quantification was not implemented. The telescope was introduced in ancient China between the Ming and Qin Dynasties in 1622 by the preacher Johann Adam Schall Von Bell, which played an important role in the brutal palace calendar battle between the old one and the new one based on observations. From solar observations, the people at that time recognized that the sunspots “drift from east to west, 14 days along the diameter, the larger one reducing the Sun’s luminosity” (Wang, Wu, and Sun, 2008). China may have a lot of sunspot drawings through telescopes, but at present, only the largest relatively comprehensive and systematic observations are gathered in our data set. This data set consists of observations from six stations, but the following three stations contribute the vast majority of data to our archives.

Qingdao Observing Station (QDOS) is the earliest in modern China to study sunspots using telescopy, but its sunspot drawings had a rather difficult history (see Table 2). Its first observation in current archives was recorded by Gao Pingzi (1888–1970) on 1st May 1925. An example observation from the same year is shown in Figure 1. He used a 16 cm equatorial telescope left by a German and installed a projector board behind the eyepiece. Then sunspots and plage were depicted by hand on a paper placed on this board, adjusting the solar diameter to 18.2 cm. One picture was drawn every day except when it was overcast.

Purple Mountain Astronomical Observatory (PMO) carried out sunspot drawings with a 20 cm aperture refractive telescope made by Zeiss company in 1934. In 1937, PMO was forced to move away from Nanjing, but in the same year, the PMO staff set up an observatory at the top of Fenghuang Mountain in the vicinity of Kunming city. Sunspot observations were made with an 8 cm refractive telescope and relative sunspot numbers were published every half year. After the victory of the war against Japan, most of PMO staff moved back to Nanjing to rebuild the observatory that was in ruins. In 1954, all sunspot drawing data in China was collected and analyzed at PMO. Since then, regular observations of daily sunspot drawings in China started, and from 1957, all sunspot drawings were required to have a solar disk image of 17.4 cm.

Yunan Astronomical Observatory (YNAO) was developed from PMO’s station in Fenghuang Mountain after the war with Japan. YNAO’s sunspot drawing tradition was inherited from PMO and its observation continuity make it best in our data set. So, YNAO contributes the largest part to our sunspots drawings archive. The sunspot drawings from YNAO were evaluated by international colleagues. Sunspots area data for each day were
Figure 1  Early original picture of sunspot drawing made by Gao Pingzi at Qingdao observatory. It shows combined sunspot evolution images that were observed on various dates: for the upper sunspot the date runs from 22 Dec 1925 to 2 Jan 1926, while for the bottom sunspot the date runs from 19 to 31 Dec 1925. For positions, they are labeled by N, S, W and E in the image.

found to have the smallest random error in the world and they had good quality to supplement the global sunspot data, and also filled time zone gaps (Baranyi et al., 2001; Balmaceda et al., 1976).

Chinese Solar-Geophysical Data (CSGD, printed journal) from 1971 to 2001 has published the sunspots’ daily relative numbers, areas and predicted smoothed numbers, which all came from the records of sunspot drawings from Chinese observatories (Yan et al., 2018). Parts of CSGD are published online by NGDC (National Geophysical Data Center), whose website is http://www.ngdc.noaa.gov/nddc.

In general, Chinese sunspot drawings contain almost complete physical information realized in that era. However, these materials were not easy to completely preserve due to the difficulties as regards long history, humidity, decay, pests damage, relocations, etc. In order to better utilize the scientific value of these data, it is necessary to extract information completely, accurately, and reliably from the Chinese historical sunspot drawings, and also make the data available through network sharing and keep the data for long-term use by digitizing the sunspot drawings so that the information from the data is preserved.

With the support of National Basic Research Program from the Ministry of Science and technology of China, the digitization of Chinese historical sunspot drawings commenced from May 2014. By the end of April 2018, we finished the digitization of sunspots drawings from six Chinese observing stations and also completed their parametric extraction. In the process of digitization we go through the following processes: the scanner selection, the scanning results evaluation of the original image, the automatic extraction of handwritten parameters, the proof of extracting parameters by computer program, and manually checking through sampling. Here, we present a part of the statistical results from the entire analysis, and the details of the process are given in the following six sections.
### Table 3  Information about observing stations and observation years in China.

| Observatory/station (abbreviation) | Observation years |
|----------------------------------|-------------------|
| Qingdao Observatory Station (QDOS) | 1925              |
| N36°04', E120°19'                 | 1947 – 1977       |
|                                  | 1980              |
|                                  | 1982 – 1989       |
|                                  | 1991 – 1992       |
|                                  | 1995 – 1996       |
|                                  | 2000 – 2009       |
|                                  | 2011 – present    |
| Purple Mountain Astronomical Observatory (PMO) | 1954 – 1963       |
| N32°03', E118°49'                | 1965 – 1980       |
|                                  | 1982              |
|                                  | 1985 – 2011       |
| Yunnan Astronomical Observatory (YNAO) | 1957 – 2016       |
| N25°01', E102°47'                | 2018 – present    |
| Sheshan Observatory Station (SSOS) | 1952 – 1964       |
| Beijing Planetarium (BJP)        | 1979 – 1982       |
| N39°54', E116°23'                | 1989 – 1999       |
| Nanjing University (NJU)         | 1986 – 2002       |
| N32°03', E118°51'                | 2004 – 2015       |

### 2. Observational Data of Chinese Sunspot Drawings

Historically, there mainly exist six observing stations for sunspot drawings in China: PMO, YNAO, QDOS, Sheshan Observing Station (SSOS), Beijing Planetarium (BJP) and Nanjing University (NJU). Table 3 lists Chinese sunspot drawings data, where the name, longitude and latitude of station, time range, number of images are recorded. The data continuity is best achieved by YANO, with a time span from 1957 to present (even if afterwards such traditional observations were no longer supported by operational funds), of 62 years. The second best is PMO with a time span of 57 years. At each individual station on sunny days, one sunspot image is drawn. In China, the first sunspot drawing was made at the QDOS in 1925. Unfortunately, the observations were soon interrupted by the Japanese invasion of Nanjing in 1937 and sunspot drawings data of nearly two years were lost. Additionally, the observations were interrupted several times due to social turmoil of that time. Regular observations of daily sunspot drawings resumed only after 1947. Since then, the observatory has been continuously performing observations. Table 3 shows that there are different observation periods for data of QDOS and PMO. However, from the perspective of complementarity of observations, there exist continuous sunspot observation data in China from 1947 to present. The main parameters of sunspot drawing telescope in all observing stations are listed in Table 4. In this paper, we mainly discuss the data processing of sunspot drawings by PMO and YNAO as examples among six observing stations. Figures 2 and 3 are sunspot drawing telescopes of PMO and YNAO, respectively.
Table 4  Main parameters of the telescopes used for drawing sunspots at the most recent individual stations.

| Observatory/station | Diameter | Focal length | Diameter of the Sun |
|---------------------|----------|--------------|---------------------|
| QDOS                | 32 cm    | 350 cm       | 17.4 cm             |
| PMO                 | 20 cm    | 350 cm       | 17.4 cm             |
| YNAO                | 12.7 cm  | 195 cm       | 17.4 cm             |
| NJU                 | 43 cm    | 217 cm       | 20 cm               |

Figure 2  PMO sunspot drawing telescope is an equatorial refractive telescope with the diameter and focal length of 20 cm and 350 cm. When it is observing, the radius of the Sun’s projection is 17.4 cm.

Figure 3  YNAO sunspot drawing telescope is an equatorial refractive telescope with the diameter and focal length of 12.7 cm and 195 cm. When it is observing, the radius of the Sun’s projection is 17.4 cm.

2.1. Processes of Sunspot Drawing

The traditional projection method was used by PMO, YNAO, etc. wherein they obtain sunspot drawings by projecting an enlarged image of the Sun onto a projection plate. At first the preprinted sunspot observation record paper is fixed on the projection plate, and the position of directions (east, west, south and north) is determined accurately. Then slowly the telescope is moved to make sure that the solar projection always overlaps with the solar limb printed on the recording paper. Finally, the specific features of sunspots are accurately drawn with a pencil. For example, according to the projection image of sunspots on the projection plate, the penumbra of the sunspots is firstly drawn with a hard pencil, and then the umbra of the sunspots is traced with a soft pencil. The western sunspots are drawn before the eastern...
ones. The larger sunspots are drawn first and then the smaller ones. After the observations, other conventional observing information is recorded: the date of observation, Beijing time (standard time of 120 degrees east longitude), international standard time (UTC) and Carrington rotation number (starting on 9 November 1853; when the prime meridian at the center of the solar disk is defined as the beginning of the first solar rotation; from then the solar rotations are numbered, the start and end dates and numbers for each solar rotation can be found in astronomical almanacs). At the same time, through the astronomical almanacs, the related parameters at the observation time are calculated, such as $P$ (the position angle between the geocentric north pole and the solar rotational north pole, in the Carrington coordinates of the solar surface, positive toward the east and negative toward the west), $B_0$ (heliographic latitude of the center of the solar disk observed at the universal time of zero on that day), $L_0$ (heliographic longitude of the center of the solar disk at zero universal time on the observation day), $L$ (heliographic longitude of the center of solar disk at the time of the observation). These data can be used to obtain heliographic coordinates, i.e. to calculate the latitude and longitude of the sunspots. Also, the area of the sunspot group can be measured: A special transparent glass plate is placed on the projection plate. Each square area in the glass plate is 1 square millimeter and the number of squares corresponding to each sunspot group is recorded (Li, Zhou, and Ji, 2016). The number of squares contained in a sunspot group can be converted into the area of sunspots by correlation calculation. Then the sunspot group is numbered and other information such as the coordinates of the sunspot group and the types of the sunspot group are recorded. Finally, the number of sunspots groups, the number of sunspots, the Wolf number on the southern hemisphere, the northern hemisphere and the whole solar surface are calculated, respectively (Rue and Loewy, 2012).

2.2. The Content of Sunspot Drawings

An example of sunspot drawings is shown in Figure 4, in which the content of the rectangular and oval boxes marked with red color are described below. The rectangular boxes are printed at fixed position on the observation paper.

Rectangular box 1 contains:
- Observing day number within the year.
- Date of observation.
- Beijing time.
- Universal Time Coordinated (UTC).

Rectangular box 2 contains:
- $P$ angle: the position angle between the geocentric north pole and the solar rotational north pole measured eastward from geocentric north.
- $B_0$: heliographic latitude of the central point of the solar disk.
- $L_0$: heliographic longitude of the central point of the solar disk.
- $L$: heliographic longitude of the center of solar disk at the time of the observation.

Rectangular box 3 contains various combinations:
- $g$: number of sunspot groups.
- $g_N$: number of sunspot groups in the northern hemisphere.
- $g_S$: number of sunspot groups in the southern hemisphere.
- $g_{NS}$: total number of sunspot groups, $g_{NS} = g_N + g_S$.
- $f$: total number of sunspots.
- $f_N$: number of sunspots in the northern hemisphere.
- $f_S$: number of sunspots in the southern hemisphere.
\( f_{\text{NS}} \): total number of sunspots, \( f_{\text{NS}} = f_N + f_S \).

R: Wolf number.
RN: Wolf number in the northern hemisphere.
RS: Wolf number in the southern hemisphere.
RNS: total Wolf number, \( R_{\text{NS}} = RN + RS \).

K: the normalized coefficient of sunspot relative number, which is related to site instrument and observer which vary with time. Here \( R_z \), published in Zurich, was used as the standard, \( K = R_z/R_y \) (\( R_y = \text{RNS} \)). After the Zurich data ceased, the standard of international sunspot relative number \( R_i \) is used as standard.

K\(_2\): It is similar to K, but for a normalized coefficient of the sunspot area.

The observer's family name.

Rectangular box 4 contains:
Weather conditions: e.g. thin cloud.
Seeing: e.g. 3. Their values range from 1 to 5, in which 1 is the best.
The information of the oval boxes are the random handwritten region information. The contents are the description of sunspots on the surface, including:
Sunspot group number.
Longitude, latitude.
Structure types of sunspot: e.g. CHI, BXI, here the McIntosh classification was used (McIntosh, 1990).
\( \alpha \): total area of the sunspot group, one in a million of area of solar disk.
\( \alpha' \): the area of the largest sunspot in the sunspot group, in units of millionth of the solar disk, referred to as the area of the largest sunspots.
r: linear distance between center of mass of the sunspot group and the center of the solar disk in units of mm.

For example in Figure 4 the oval handwritten area for the No. 63 sunspot group gives latitude and longitude of \(-16.0^\circ\) and \(+44.0^\circ\), sunspot group structure type CSI group area of \( \alpha = 1.3 \), the area of the largest sunspot \( \alpha' = 1.1 \), linear distance between center of mass of sunspot group and the center of the solar disk \( r = 61 \).

3. Digital Processing

For the digital processing, we visited the World Data Center in Brussels, Belgium, and entered collaboration with Dr. Frederic Clette who was converting Europe’s historical sunspot drawings into digital files. Referring to their standard suggestions and many scanning trials, we established a few regulations for the guidance of ever-repeated scanning work.

3.1. Scanner Selection

For the selection of the scanner, there are three basic requirements. 1) Resolution: the description of the smallest spot on the sunspot drawing is about the size of the pencil tip (0.05 mm). To clearly scan such a point, at least 2 pixels are needed, which requires the resolution of the scanner to be > 1000 ppi (pixels per inch). 2) Gray level: in order to effectively distinguish the umbra from the penumbra (due to the long storage time of the early drawings and light color), the scanner with the maximum gray level is required. 3) Width: A3. After careful investigation, the Zhongjing 1960XL scanner that met the above requirements was chosen.
Figure 4  An example of sunspot drawings observed by YNAO observed on 18 March 2013 01:00 UTC. Here printed part information is recorded in the rectangle region, while the handwritten part information relating specific and individual sunspot is recorded in the oval boxes. The orientation (N, S, W and E) and angle-scale (degree) are marked on the circumference.

3.2. Scanning Rules

The rules of scanning originates from two aspects: visions and qualities. The sources of different time (year) and observers are selected and put into the scanner. The main operation is to adjust the brightness and contrast, for example, when the brightness of gray scale of
Figure 5  The position of a sunspot drawing in the scanner. The upper and left edges of the image coincide strictly with those of the scanner, the north of the image points up, while the west is to the left. At the same time, try to the best to keep the paper flat and close to the printer.

image is too bright or too dark, then drag the brightness slider to change it. If the brightness is too high, the image will look white. It will be too dark if the brightness is too low. The brightness of the image should be set to moderate by the brightness slider. For other parameters, the local changes can be made according to a similar adjustment method. The team members carried out visual inspection until everyone was satisfied, and then formulated the technical parameters and operational rules for the scanning of sunspots. Quality assessment was carried out by comparing the relative number of sunspots between digital and original data and repeatedly verified. The preparation ends when requirements (relative error < 5%) are satisfied. After the scanner is adjusted according to the above rules, the specific rules of scanning process are as follows: 1) Prepare a piece of pre-scanned original sunspot drawing, check the conditions of the folding, wrinkle and dust on the surface, and carefully clean the paper surface and flatten it. 2) Put the original image face down into the scanner, adjust the position, and align the scanned paper with the corner of the scanner. Make sure that the north–south direction of the observation paper is perpendicular to the scanning direction, as shown in Figure 5. 3) In order to save all the details of the original observation, the 24-bit BMP (Bit Map Picture) color image format was chosen to save the image, that resulted in the size of a single image about 90 MB. Although this data format is more than 10 times larger than the commonly used JPG (Joint Photo Graphic) format, the original observations can be recorded truly without compression distortion (Figure 4), which allowed us to work with great precision to analyze each of the images.

3.3. File Naming and Storage

Since the data was produced from multiple stations, the file name and storage directory of all observational data that were scanned are specified for convenience of their database query. The file name is consistent with the basic rules of international astronomical naming, and
the unified form is

<observatory>_<sd>_<year>_ <month>_ <date>_ <hour>_ <minute>_ <observer>_.bmp

The observatory refers to the source of observational data, the abbreviation of the full name of the Chinese observatory refers to the first column of Table 3, Sd refers to the type of observational data which indicates sunspot drawing. The format of observation time is year, month, date, hour and minute (UTC). For example, for the sunspot drawing made by a YNAO observer whose last name is Ye at 5:40 on 18 February 1957, the file name after digitization is as follows:

Ynao_sd_19570218_0540_ye.bmp

After scanning, the file is stored on the server, and the folder is set as follows:

/1957/02/YNAO/

4. Automated Extraction of Parameters

Based on the scanned image and convolutional neural network (CNN), an identification software is developed to determine the sunspot drawing parameters automatically, which is further described in Section 2.2 of the automatic identification of the content in the rectangular (printed parts) and oval box (handwritten parts). Figure 6 shows the interface of identification software, wherein the content on the left is the recognition results shown by the middle image, which includes the whole content of rectangular and oval box. For a detailed introduction of this software, we refer the reader to a former paper (Zheng et al., 2016).

4.1. Parameter Record Format

The identified parameters are stored in the “.txt” document format (on the right of Figure 7) and the “.csv” table format (see Tables 5, 6), which are available to users for further processing them as required. In the TXT document format, the printed part of the data content begins with ‘#’, ‘,’ is a separator, ‘;’ means end, ‘?’ indicates that the information is unknown (or, due to the longer time the handwriting is blurred and cannot be recognized; or there is no original information), there is no space between the separator. The handwritten part of the data content begins with ‘∗’, ‘,’ is separator, ‘;’ is end, ‘?’ means the information is unknown (the reason is the same as above), as shown in Figure 7 on the right. An example of records is shown in Figure 7, wherein the information is labeled by red/green color boxes that were recorded as red/green rectangle labeled in documents, respectively.

5. Accuracy Verifications of Extraction Parameter

Due to the large number of sunspot drawings, YNAO alone has 15 752 sunspot drawings, and there are 1 051 422 records of information in rectangular and oval boxes that contained in images. In order to verify the original data and ensure the reliability of digital data, the following verification methods are used to check the data, to finally carry out the accuracy test.
Table 5  Example of the CSV table printed parts.

| Num. | Date     | B0 (UTC) | Day | L0 | L | gN | gS | gNS | fN | fS | fNS | RN | RS | RNS | K | R | Vis. | K2 |
|------|----------|----------|-----|----|---|----|----|-----|----|----|-----|----|----|-----|---|---|-----|---|
| 286  | 19841110 | 850      | 50  | 0.035 | 22.67 | 3.37 | 276.24 | 275.78 | 2 | 0 | 2 | 14 | 0 | 14 | 34 | 0 | 34 | 0.65 | 22 | 3 | 0.99 |
| 287  | 19841111 | 1020     | 220 | 0.097 | 22.43 | 3.26 | 263.08 | 261.77 | 2 | 1 | 3 | 14 | 3 | 17 | 34 | 13 | 47 | 0.65 | 31 | 3 | 0.99 |
| 288  | 19841112 | 900      | 100 | 0.042 | 22.18 | 3.14 | 249.87 | 249.32 | 1 | 1 | 2 | 10 | 1 | 11 | 20 | 11 | 31 | 0.65 | 20 | 3 | 1.08 |
| 289  | 19841113 | 850      | 50  | 0.035 | 21.92 | 3.03 | 236.69 | 236.23 | 1 | 0 | 1 | 8  | 0 | 8  | 18 | 0  | 18 | 0.65 | 12 | 3 | 1.08 |
| 290  | 19841114 | 840      | 40  | 0.028 | 21.65 | 2.91 | 223.5  | 223.13 | 1 | 0 | 1 | 4  | 0 | 4  | 14 | 0  | 14 | 0.65 | 9  | 2 | 1.08 |
| 291  | 19841115 | 900      | 100 | 0.042 | 21.38 | 2.8  | 210.32 | 209.77 | 0 | 1 | 1 | 0  | 3  | 3  | 0  | 13 | 13 | 0.65 | 8  | 2 | 1.08 |
| 292  | 19841116 | 940      | 140 | 0.07  | 21.10 | 2.68 | 197.14 | 196.22 | 0 | 1 | 1 | 0  | 2  | 2  | 0  | 12 | 12 | 0.65 | 8  | 3 | 1.08 |
| 293  | 19841117 | 905      | 105 | 0.045 | 20.81 | 2.56 | 183.95 | 183.35 | 0 | 1 | 1 | 0  | 2  | 2  | 0  | 12 | 12 | 0.65 | 8  | 2 | 1.08 |
| 294  | 19841118 | 835      | 35  | 0.024 | 20.52 | 2.44 | 170.77 | 170.44 | 1 | 0 | 1 | 4  | 0 | 4  | 14 | 0  | 14 | 0.65 | 9  | 2 | 1.08 |
| 295  | 19841119 | 830      | 30  | 0.021 | 20.22 | 2.32 | 157.59 | 157.31 | 1 | 0 | 1 | 5  | 0 | 5  | 15 | 0  | 15 | 0.65 | 10 | 2 | 0.99 |
| 296  | 19841120 | 810      | 10  | 0.007 | 19.91 | 2.2  | 144.41 | 144.32 | 1 | 1 | 2 | 6  | 1 | 7  | 16 | 11 | 27 | 0.65 | 18 | 3 | 0.99 |
| 297  | 19841121 | 945      | 145 | 0.073 | 19.59 | 2.08 | 131.23 | 130.26 | 1 | 1 | 2 | 13 | 2 | 15 | 23 | 12 | 35 | 0.64 | 22 | 3 | 1.01 |
| 298  | 19841122 | 900      | 100 | 0.042 | 19.26 | 1.95 | 118.05 | 117.5 | 1 | 2 | 3 | 11 | 4 | 15 | 21 | 24 | 45 | 0.64 | 29 | 3 | 1.01 |
| 299  | 19841123 | 840      | 40  | 0.028 | 18.93 | 1.83 | 104.87 | 104.5 | 1 | 2 | 3 | 7  | 15 | 22 | 17 | 35 | 52 | 0.64 | 33 | 3 | 1.01 |
| 300  | 19841124 | 940      | 140 | 0.07  | 18.59 | 1.71 | 91.69  | 90.77 | 1 | 2 | 3 | 6  | 18 | 24 | 16 | 38 | 54 | 0.64 | 35 | 3 | 1.01 |
| 301  | 19841125 | 820      | 20  | 0.014 | 18.25 | 1.58 | 78.51  | 78.33 | 1 | 2 | 3 | 2  | 34 | 36 | 12 | 54 | 66 | 0.64 | 42 | 3 | 1.01 |
Figure 6  Sunspot drawing identification software. The left parts show the identifiable information that use automatic recognition method; the middle part shows the original image and observed information; while the right parts give the some necessary operations such as archiving and proofreading.

Figure 7  TXT record format and its corresponding image content. The red part region labeled in TXT corresponds to the printed information within the red rectangles in the original image; while the green part region labeled in TXT is the information identified from the handwritten characters in the original sunspot drawings.

5.1. Check the Accuracy of Values by Data Type

Each sunspot drawing contains different types of sunspot data, and the length (number of digits) of different types of values are different, but the data format of the same type of data
Table 6  Example of the CSV table handwritten parts.

| Date       | Order | Latitude | Longitude | Type | α   | α'  | r   | N  |
|------------|-------|----------|-----------|------|-----|-----|-----|----|
| 19580107   | 43    | 21       | 49        | C    | 0.4 | 0.4 | 70  | 4  |
| 19580107   | 44    | 16       | 48        | C    | 1.4 | 0.7 | 68  | 5  |
| 19580107   | 45    | -19      | 40        | A    | 0.2 | 0.2 | 58  | 2  |
| 19580107   | 47    | -22      | 33        | C    | 0.8 | 0.7 | 52  | 5  |
| 19580107   | 48    | 4        | 14        | C    | 0.6 | 0.5 | 24  | 4  |
| 19580107   | 50    | -17      | 1         | C    | 1   | 0.9 | 20  | 3  |
| 19580107   | 51    | 11       | -2        | F    | 9   | 6   | 22  | 64 |
| 19580107   | 52    | -28      | -38       | C    | 1.2 | 1   | 60  | 4  |
| 19580107   | 53    | -17      | -40       | D    | 3.6 | 2.3 | 55  | 9  |
| 19580107   | 54    | -11      | -56       | E    | 10  | 4   | 70  | 21 |
| 19580107   | 55    | -16      | 5         | B    | 0.1 | 0.1 | 21  | 4  |
| 19580107   | 56    | -38      | -13       | C    | 0.6 | 0.4 | 52  | 6  |
| 19580107   | 57    | 14       | -31       | A    | 0.1 | 0.1 | 50  | 2  |
| 19580107   | 58    | 26       | -68       | J    | 0.8 | 0.8 | 82  | 1  |

is uniform. For example, in Table 5, number, date of observation, Beijing time, UTC, gN, gS, gNS, fN, fS, fNS and so on are positive integers, but P, B0, L0 and so on are decimals. The numeric type of the data can be determined by verifying the number of decimal places in the data, and checking whether the number of decimal places is consistent with the length of the standard numerical values.

5.2. Abnormal Value Elimination

Most of the sunspot drawing recorded by YNAO are limited to a specific range, e.g., the year of observation is 1958 – 2015; the monthly range is from January to December; for the range of days, the longer months are from 1 to 31 and the shorter months are from 1 to 30; February is from 1 to 29/28 for leap/common years; P angle varies from $-26.31$ to 26.31; $B_0$ varies from $-7.23$ to 7.23; seeing is an integer between 1 and 5. Different types of data are sorted separately, and then verified to determine whether the data is valid and its value is within the normal range. If not, the data is then manually compared with the data recorded in the sunspot drawings of that day.

5.3. The Correlation Verification

Correlation verification consists of two parts, one of which is the conversion and comparison of relevant data in the printed part. The printed part of the sunspot drawings recorded by YNAO contains information such as Beijing time, UTC, gN, gS, gNS, fN, fS, fNS, etc. There exists a certain conversion regularity, e.g., the time difference between Beijing time and UTC is about 8 hours; the number of sunspots on the visible solar surface is the sum of the number of sunspots on the northern and southern hemispheres, namely, the sum of gN and gS is equal to gNS. This rule can also be applied to fN, fS, fNS, and RN, RS, RNS. In addition, there are the following correlations between the number of sunspots, the number of sunspot groups, and the Wolf number:
\[ RN = 10gN + fN, \]
\[ RS = 10gS + fS, \]
\[ RNS = 10gNS + fNS \]
\[ R = K(10g + f), \text{ where } R = RNS, f = fNS. \]

Another verification step is to compare the data recorded in the printed and handwritten parts. For example, the printed part recorded the total number of sunspot groups and sunspots observed on that day, while the handwritten part recorded the specific information of each sunspot group, including the number of sunspots in each sunspot group. In this way, the total number of sunspot groups on a given day can be obtained by counting the number of sunspot groups recorded in the handwritten part. When the number of sunspots labeled in each sunspot group is added, the total number of sunspots on that day can be obtained. Finally, it is compared with the values of \( gNS \) and \( fNS \), respectively. Also, the method can be used to confirm if there is a sunspot group that has not been identified in the handwritten part.

5.4. Verification of the Relevant Laws

In 1861, Carrington discovered what we now know as Spörer’s law, or the butterfly diagram (Hopkins, 1976; Carrington, 1969), which was investigated in detail by Spörer. During the solar cycle, sunspots are distributed within a \( \pm 45^\circ \) latitude range, of which most sunspots appear on both sides of the equator within latitudes of \( 15^\circ - 20^\circ \) that are parallel to the equator; however, there are very few that appear on both sides of the equator’s \( 8^\circ \) range. At the beginning of the sunspot cycle, sunspots mostly appear on the surface within \( 30^\circ \) to \( 45^\circ \) latitude. As the cycle progresses, sunspots appear in lower and lower latitudes. When the average latitude equals \( 15^\circ \), the sunspot number reaches its maximum. The average latitude of sunspots will continue to decrease, at the end of the cycle the sunspots appear in about \( 7^\circ \). Then the new cycle of sunspots begins to appear at the relatively high latitudes (Phillips, 1995). For the processed sunspots, the validation is checked as follows: if the latitude of the sunspot group is located between \( -45^\circ \) to \( 45^\circ \), the above rules are followed. If not, proofread them manually. At the same time, in the solar cycle, some other parameters of the same sunspot group, such as the coordinates of longitude and latitude, area and number of sunspots, also change with the formation and decay of the sunspot group. The size and number of sunspots in an individual sunspot group usually increase from smaller to larger and then they reverse, which rule can also be used to verify the accuracy of the data. Furthermore, sunspots generally appear in the eastern hemisphere and disappear in the west. In this way, the longitude variation of the sunspot group can be verified, then it is checked whether it conforms to Spörer’s law in each sunspot cycle.

5.5. Data Test

By random sampling, a total of 160 sunspot drawings (~1%) were selected to test human recognition and input errors. The specific test scheme is as follows: Setting up a group of 10 team members, then divide them into five teams. Each team is equipped with two computers, one used to identify the original image information and the other to check the digitized data. 160 images are selected randomly from the data set of sunspot drawings, in this sample the sunspot related information was 918 (namely, the total lines of Table 5) and the valid data was 10864. According to the test results, the number of data errors in the fixed region (printed parts in Table 5) was 6, the number of errors in the relevant information of sunspots
Table 7  Test results.

| Error type | Error year | Number of errors | Sum errors | Total errors |
|------------|------------|-----------------|------------|--------------|
| B0         | 19610630   | 1               | 3          | 6 printed parts |
|            | 19631105   | 1               |            |              |
|            | 19770803   | 1               |            |              |
| P          | 19691217   | 1               | 3          |              |
|            | 19750621   | 1               |            |              |
|            | 19830528   | 1               |            |              |
| α          | 19591201   | 1               | 3          | 29 handwritten parts |
|            | 19681109   | 1               |            |              |
|            | 19830228   | 1               |            |              |
| α'         | 19680331   | 1               | 3          |              |
|            | 19681109   | 2               |            |              |
| r          | 19680331   | 2               | 11         |              |
|            | 19681109   | 1               |            |              |
|            | 19741010   | 1               |            |              |
|            | 19781208   | 1               |            |              |
|            | 19810128   | 2               |            |              |
|            | 19890707   | 1               |            |              |
|            | 19890624   | 2               |            |              |
|            | 19920225   | 2               |            |              |
| Lati.      | 19680331   | 1               | 4          |              |
|            | 19681109   | 1               |            |              |
|            | 19810421   | 1               |            |              |
|            | 19810503   | 1               |            |              |
|            | 19931226   | 1               |            |              |
| Long.      | 19800611   | 1               | 2          |              |
|            | 19890409   | 1               |            |              |
| Num.       | 19730505   | 1               | 4          |              |
|            | 19780411   | 1               |            |              |
|            | 19830528   | 1               |            |              |
|            | 20010517   | 1               |            |              |
| Group type | 19890624   | 1               | 2          |              |
|            | 20040909   | 1               |            |              |

was 29 (handwritten parts as Table 6), and the total number of data errors was 35. In these 160 data samples, the error rate of the fixed area is 3.75% (6/160); while for a line with information as regards a cluster of sunspots the error rate is 3.15% (29/918). According to the calculation of valid data, the error rate for this sample with 160 sunspot drawings was 0.32% (35/10864). The error data is shown in Table 7.

6. Data Summary and Analysis

Digitization, parameter extraction and quantitative statistics for decades of sunspot drawings in China are performed based on scanning mode of human–computer interaction and
Table 8  Statistical results of digitized sunspot drawings.

| Obs.  | Years               | Num. of sunspot drawings | Printed part (records) | Handwritten part (records) | Total num. of information |
|-------|---------------------|--------------------------|------------------------|---------------------------|--------------------------|
| YNAO  | 1957 – 2015         | 15752                    | 15752                  | 89262                     | 1051422                  |
| BJP   | 1979 – 1982         | 3764                     | 3764                   | 28647                     | 170704                   |
|       | 1989 – 1999         |                          |                        |                           |                          |
| PMO   | 1954 – 1963         | 12508                    | 12508                  | 57570                     | 626765                   |
|       | 1965 – 1980         |                          |                        |                           |                          |
|       | 1982                |                          |                        |                           |                          |
|       | 1985 – 2011         |                          |                        |                           |                          |
| QDOS  | 1925                | 12634                    | 12634                  | 48424                     | 395047                   |
|       | 1947 – 1977         |                          |                        |                           |                          |
|       | 1980                |                          |                        |                           |                          |
|       | 1982 – 1989         |                          |                        |                           |                          |
|       | 1991 – 1992         |                          |                        |                           |                          |
|       | 1995 – 1996         |                          |                        |                           |                          |
|       | 2000 – 2009         |                          |                        |                           |                          |
|       | 2011 – 2016         |                          |                        |                           |                          |
| SSOS  | 1952 – 1964         | 2415                     | 2415                   | 11586                     | 59471                    |
Method of CNN handwritten character recognition. The results shown in Table 8 indicate the number of images is 47,073, for which the digitized results are stored in the format described in Section 4.1. Among them, the number of valid records included in the printed part is 47,073, and in the handwritten part it is 235,489. Statistical distribution is carried out on the digital data of PMO and YNO with a long observation time. Temporal evolutions of full-disk sunspot number based on the digitized sunspot drawings observed by YNAO (upper panel) and PMO (bottom panel) are shown in Figure 8, where the $X$-coordinate is year and the $Y$-coordinate indicates the total number of sunspots directly observed by sunspot drawing telescope on a given day. The red solid curve is the 180 days smoothed curve superposed on the original data. The sunspot drawings from YNAO and PMO together contributed the largest part to our data archives. At present, the pure full-disk sunspots counts, which do not include the group numbers, are accurate statistics for these data archives (the sunspot relative number $R$ can be provided after the cross-calibration with the international sunspot relative number will have been finished). Figure 9 shows the cyclic variation of sunspots obtained from the digital data of PMO and YNAO, where the $X$-coordinate represents time (year) and the $Y$-coordinate mean latitude. This result provides the pattern of the butterfly diagram, which shows the variation of sunspot latitudes in the northern and southern hemispheres over the past 58 years. It is evident that the length of the average solar cycle is 11 years, and the variation of the sunspot latitude is consistent with Spörer’s law. For example, at the beginning of the 20th solar cycle (October 1964–June 1976), sunspots of the new cycle appeared to be at higher latitudes, and sunspots of the old cycle appeared at lower latitudes.

The statistics of the data continuity of each station is carried out, and the results are shown in Figure 10. Among them, PMO and YNAO have the longest observation time.
Figure 9  Butterfly diagram according to heliographic latitudes recorded in the digitized sunspot drawings obtained by YNAO (upper panel) and PMO (bottom panel). The PMO sunspot drawing observers had the habit of only recording integer values for the sunspot location, resulting in $1^\circ$ resolution. The YNAO observers tended to record decimal values before 1982, but mainly rounded to $0.5^\circ$.

Part of the monthly data is missing. We have counted the periods without sunspot drawings for more than seven days, and part of these results are given in Table 9. The YNAO has an interval of 100 times over seven days without images, while for PMO we have 283 times.

Sunspot drawings originating from various regions of China are digitized, and the corresponding software is developed to recognize and extract the valuable parameters; correctness verifications are also carried out at the same time. Additionally, according to the types and characteristics of data recorded in the image, a detailed data consolidation scheme and different data verification methods are developed. Based on this verification method, the accuracy of parameters is checked again. Finally, the database of sunspot
Figure 10  Data continuity statistics of sunspot drawings at different observatories. The different colors of patches correspond to individual observatories labeled on the Y-axis: violet, blue, dark khaki, yellow, brown and green indicate YNAO, NJU, SSOS, QDOS, PMO and BJP. The observing years are plotted on the X-axis, which can show the time coverage of data that was obtained by an observatory. At the same time, the segmented observation periods for individual observations are labeled under the corresponding color patch.

Table 9  Example list of no observation of more than seven days duration.

| Obs. | Years | Date       |
|------|-------|------------|
| YNAO | 1958  | 0208 – 0214 |
|      |       | 1122 – 1201 |
|      | 1959  | 0221 – 0301 |
|      |       | 0524 – 0531 |
|      | 1960  | 0523 – 0613 |
|      |       | 0713 – 0721 |
|      | 1963  | 0101 – 0109 |
|      |       | 0608 – 0614 |
|      |       | 1016 – 1022 |
|      |       | 1226 – 0101 |
| PMO  | 1956  | 0116 – 0123 |
|      |       | 0127 – 0204 |
|      |       | 0320 – 0402 |
|      |       | 0605 – 0612 |
|      |       | 0731 – 0807 |
|      |       | 0824 – 0831 |
|      |       | 0913 – 0925 |
|      | 1957  | 0107 – 0118 |
|      |       | 0131 – 0213 |
|      |       | 0420 – 0503 |
|      |       | 0727 – 0807 |
|      |       | 1123 – 1130 |
|      |       | 1206 – 1215 |
drawings in China and their parameters is established. A preliminary analysis of the data shows that it is consistent with the latitudinal distribution pattern of sunspots. The work of digitizing Chinese sunspot drawings and extracting their parameters not only makes Chinese historical sunspot observations permanently preserved, but also makes the data have continuity, integrity, complementarity and usability. These four aspects fill the gaps in the digital observational data of sunspots in this geographical region, and provide more detailed and richer data for the study of long-term sunspot evolution. The database is available on the web site: http://sun.bao.ac.cn/SHDA_data/. At present, YNAO and QDOS continue to produce sunspot drawings, which can provide the continuity of data for the ground-based observation of sunspots for the Chinese timezone. The quality assessment of the sunspot drawings in China is under way and corresponding results will be published. A web-based English version query interface will be released soon (database is already accessible by the web site: http://sun.bao.ac.cn/SHDA_data/). In the next step, we will pay more attention to these extracted parameters and work more accurately on the basis of combining original sunspot drawings and produce a newer version of the parameter set.

7. Data Policy

The images and data from the Archive of Chinese Historical Sunspot Drawings (ACHSD) can be freely downloaded as public data. However, any public use, web based or paper publication, of those data must include an explicit credit to the source: (ACHSD data/image, National Astronomical Observatories, CAS, Beijing, China).

Acknowledgements We thank the reviewer for valuable suggestions and constructive criticism, which improved the clarity of the article. The work was funded by National Science Foundation of China Grant Nos.: u1531247 and 1142790111427901, the 13th Five-year Informatization Plan of Chinese Academy of Sciences, Grant No. XXH13505-04 and the special foundation work of the ministry of science and technology of the of China Grant No.: 2014fy120300. We would like to thank our predecessors who have been engaged in the observation of sunspot in China over the past decades.

Disclosure of Potential Conflicts of Interest The authors declare that they have no conflicts of interest.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Balmaceda, L.A., Solanki, S.K., Krivova, N.A., Foster, S.: 1976, J. Geophys. Res. 114, A07104. DOI. ADS.
Baranyi, T., Győri, L., Ludmány, A.: 2015 In: IAU General Assembly, 2257669. ADS.
Baranyi, T., Győri, L., Ludmány, A., Coffey, H.E.: 2001, Mon. Not. Roy. Astron. Soc. 323, 223. DOI. ADS.
Carrington, R.C.: 1969, Observations of the spots on the Sun from November 9, 1853, to March 24, 1861, made at Redhill. [M]. Readex Microprint.
Clette, F., Svalgaard, L., Vaquero, J.M., Cliver, E.W.: 2014, Space Sci. Rev. 186, 35. DOI. ADS.
Hopkins, J.: 1976, Glossary of Astronomy and Astrophysics, University of Chicago Press, Chicago. 174 p., ADS.
Howard, R.F.: 1991, Special. Pub. 50, NASA, Washington, D.C., 451.
Howard, R.F., Gilman, P.L., Gilman, P.A.: 1984, Astrophys. J. 283, 373. DOI. ADS.
Leo, R.: 2011, Early Astronomy and the Beginnings of a Mathematical Science, NRICH (University of Cambridge)
Li, Y.: 2015, Acta Astron. Sin. 56, 528. ADS.
Li, R.Y., Zhou, T.H., Ji, K.F.: 2016, *Acta Astron. Sin.* **57**, 292. DOI. ADS.

McIntosh, P.S.: 1990, *Solar Phys.* **125**, 251. DOI. ADS.

Phillips, K.J.H.: 1995, *Guide to the Sun*. Cambridge University Press, Cambridge. ADS.

Ravindra, B., Priya, T.G., Amareswari, K., Priyal, M., Nazia, A.A., Banerjee, D.: 2013, *Astron. Astrophys.* **550**, 19. DOI. ADS.

Rue, W.D., Loewy, B.: 2012, *Phil. Trans. Roy. Soc. London* **159**, 1.

Sivaraman, K.R., Gupta, S.S., Howard, R.F.: 1993, *Solar Phys.* **146**, 27. DOI. ADS.

Tang, J.: 2014, *Acta Astron. Sin.* **55**, 137. ADS.

Tang, J.: 2015, *Chin. Astron. Astrophys.* **39**, 45. DOI. ADS.

Ulrich, R.K., Webster, L.S., Varadi, F., Javaraiah, J., Lefebvre, S., Gilman, P.: 2004 In: AGU Fall Meeting Abstracts, SH52A-03. ADS.

Vanlommel, P., Cugnon, P., van der Linden, R.A.M., Berghmans, D., Clette, F.: 2004, *Solar Phys.* **224**, 113. DOI. ADS.

Wang, G.C., Wu, Y.H., Sun, X.C.: 2008, *Stud. Hist. Nat. Sci.* **27**, 309.

Willis, D.M., Henwood, R., Wild, M.N., Coffey, H.E., Denig, W.F., Erwin, E.H., Hoyt, D.V.: 2013, *Solar Phys.* **288**, 141. DOI. ADS.

Wittmann, A.D., Xu, Z.T.: 1987, *Astron. Astrophys. Suppl. Ser.* **70**, 83. ADS.

Yan, Y., Du, Z.L., Wang, H.N., He, H., Guo, J., Huang, X., Zhu, X.S., Dai, X.H., Lin, G.H.: 2018, *Proc. IAU* **13**(S340), 71. DOI.

Zheng, S., Zeng, X., Lin, G., Zhao, C., Feng, Y., Tao, J., Zhu, D., Xiong, L.: 2016, *New Astron.* **45**, 54. DOI. ADS.