Yunnan Observatory Hand-Drawn Sunspot Data System: Postprint

Authors: Zhu Gaofei, Zheng Sheng, Lin Ganghua, Zeng Xiangyun, Feng Yongli, Tao Jinping, Shu Yao

Date: 2018-05-23T00:00:00+00:00

Abstract

Yunnan Observatories, Chinese Academy of Sciences has accumulated nearly 16,000 sheets of hand-drawn sunspot observation data over the years, necessitating the establishment of a comprehensive query and statistical system for scientific management and statistical analysis of these massive datasets. The digitized data includes nearly 90,000 sunspot records and over 1 million valid data entries. The system provides a platform for data management, retrieval, and statistical analysis. Using the data management system, long-term statistical analysis can be performed on sunspot-related parameters.

Full Text

Hand-drawn Sunspot Data System of Yunnan Observatories

Zhu Gaofei[1,2], Zheng Sheng[1], Lin Ganghua[2], Zeng Xiangyun[1,2], Feng Yongli[3], Tao Jinping[3], Shu Yao[1]

[1]College of Science, China Three Gorges University, Yichang 443002, China
[2]Key Laboratory of Solar Activity, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China
[3]Yunnan Observatories, Chinese Academy of Sciences, Kunming 650011, China

Abstract: The Yunnan Observatories of the Chinese Academy of Sciences has accumulated nearly 16,000 hand-drawn sunspot observations over many years. To scientifically manage and analyze these massive datasets, it is necessary to establish a comprehensive query and statistical analysis system. The digitized data includes nearly 90,000 sunspot records and over one million valid data entries. The system provides a platform for data management, retrieval, and statistical analysis, enabling long-term statistical analysis of sunspot-related parameters.

Keywords: Sunspot; Data management; Sunspot information retrieval; Sunspot statistical analysis

Among various solar activities, sunspots are the most conspicuous and easily observable phenomenon. Telescopic observations of sunspots began intermittently in 1610, with regular daily observations starting in 1818, yielding relatively reliable sunspot data. Early sunspot observations were typically hand-drawn on paper, resulting in a large volume of paper-based records. Due to the fragility of paper and the difficulty of information retrieval, data integration and sharing among observatories have been hindered. Consequently, converting these paper images into digital information storage is essential for facilitating analysis and access.

China' s solar physics observation data possesses a time zone advantage, making it internationally valuable and scarce. The hand-drawn sunspot images from Yunnan Observatories contain rich information and represent invaluable historical data for long-period solar activity research. The sunspot area data observed by Yunnan Observatories from 1981 to 1992 has been used multiple times to supplement and improve global sunspot datasets. Recognizing the unique value of China' s hand-drawn sunspot observations, a systematic scanning, organization, and digital image storage project for nearly a century' s worth of historical hand-drawn sunspot data from Yunnan Observatories and Purple Mountain Observatory was initiated in 2014, led by the Huairou Solar Observing Station of the National Astronomical Observatories. To maximize the scientific value of these historical observations for studying long-term variations in solar activity, it is necessary to accurately and reliably extract sunspot information from the scanned digital images and enable online sharing. Image analysis techniques are employed to segment handwritten characters, and deep learning methods are used for character recognition to achieve automatic extraction of hand-drawn sunspot information, thereby digitizing the sunspot drawings.

Systematic management of the original digitized sunspot data enables database retrieval of sunspot information and statistical analysis of various sunspot variations, providing convenience and ease of use. Most importantly, it allows for the aggregation of observational data from various observatories and effective data organization to facilitate comprehensive and in-depth studies of sunspot activity and its effects on Earth. Therefore, systematic management and analysis of the digitized hand-drawn sunspot data is particularly important. This paper describes the construction of the Yunnan Observatories Hand-drawn Sunspot Data System using the MVC pattern combined with EF (Entity Framework) technology, enabling integrated data management and clear, intuitive statistical analysis.

1 Data Acquisition

The digital scanning of hand-drawn sunspot images from Yunnan Observatories for the period 1958–2015 has been completed. Each hand-drawn sunspot image

contains substantial information of diverse types, primarily including routine observation records and sunspot information, as illustrated in Figure 1 [Figure 1: see original paper].

The rectangular boxes in Figure 1 contain the following information:

Box 1: Rotation number (the numbering system where the moment when the prime meridian reached the solar central meridian on November 9, 1853, is defined as the start of the first solar rotation; rotation numbers and start dates for each year can be found in astronomical almanacs), observation date, Beijing time (East Longitude 120° standard time), and Coordinated Universal Time (UTC).

Box 2: Position angle P (the azimuth angle of the solar rotation axis north pole measured from the solar north point), B and L (the latitude and longitude of the solar central meridian at 0h UT on the observation day), and L (representing the longitude of the solar central meridian at the observation time).

Box 3: Number of sunspot groups in the northern hemisphere (g_N), number of sunspot groups in the southern hemisphere (g_S), total number of sunspot groups (g_NS), number of individual sunspots in the northern hemisphere (f_N), number of individual sunspots in the southern hemisphere (f_S), total number of individual sunspots (f_NS), Wolf number for the northern hemisphere (R_N), Wolf number for the southern hemisphere (R_S), total Wolf number (R_NS), observatory factor (k), and Wolf number (R).

Box 4: Weather conditions, visibility, and remarks.

Ellipse: Sunspot group number, longitude, latitude, sunspot group structure type, total area of individual sunspot group, area of the largest sunspot in the group (in millionths of the solar disk area, abbreviated as largest spot area), and straight-line distance from the sunspot group centroid to the center of the solar projection disk (in millimeters, abbreviated as radius).

Computer-assisted manual processing is used to digitize the above information from the scanned hand-drawn sunspot images. Reference [11] details the application of deep learning methods for digitizing information from hand-drawn sunspot images. Through image segmentation, handwritten characters in the sunspot drawings are separated and then recognized using deep learning methods. The digitized information is stored in two data tables, with the record formats shown in Table 1 and Table 2 , containing nearly 90,000 sunspot information records and over one million valid data entries.

2 System Design

2.1 Function Design

The completed system not only provides retrieval and download of sunspot observation data but also enables multi-faceted statistical analysis of the data, allowing researchers to intuitively understand variation trends in sunspot-related

parameters. The system is divided into three major functional modules with multiple sub-modules, as shown in Figure 2 [Figure 2: see original paper]: Sunspot Information Database, Sunspot Information Query, and Sunspot Information Statistics.

2.2 Architecture Design

The system is built using the MVC (Model-View-Controller) design pattern, which separates business logic, data, and interface display. This organizational structure concentrates business logic into a single component, allowing interface improvements and customization of user interaction without rewriting the business logic. The system architecture is illustrated in Figure 3 [Figure 3: see original paper].

2.2.1 Model Class Design The model class design provides convenient and accessible entity classes for controller operations and view invocation. The system includes two primary model classes and multiple auxiliary view classes:

(1) Basic Information Class: This class primarily maps one-to-one data information from the hand-drawn sunspot images, such as date, number, standard time, international time, weather conditions, visibility, and remarks.

(2) Sunspot Information Class: This class maps various sunspot parameter information for each sunspot (group) in the hand-drawn images, including sunspot number, total area of individual sunspot group, area of the largest sunspot, longitude, latitude, and radius.

2.2.2 Controller Design Controllers handle user requests and invoke appropriate views for display. The system includes three main controllers and multiple auxiliary controllers:

(1) Sunspot Main Controller: This controller manages the display of various information from the hand-drawn sunspot images in the main database, including list displays of drawing information and detailed displays of sunspot information.

(2) Sunspot Query Controller: This controller processes data retrieval tasks based on different parameters according to user requests, such as retrieving corresponding sunspot information by year, sunspot group number, total area of individual sunspot group, area of the largest sunspot, longitude, latitude, and other criteria, then feeding back results through appropriate views.

(3) Sunspot Statistics Controller: This controller functions as an analytical processing unit, primarily responsible for statistically analyzing the total area of individual sunspot groups, area of the largest sunspots, radius, and variations in total sunspot group area and quantity, presenting results intuitively to users through charts.

2.2.3 View Design View design, also known as UI interface design, is primarily responsible for receiving data from controllers and presenting it to users in a friendly manner. The system includes a main view, shared views, partial views, and various display views corresponding to different controllers.

3 System Functions

3.1 Sunspot Information Database

The Sunspot Information Database displays all digitized information from Yunnan Observatories' hand-drawn sunspot images, navigated by year, presenting all observed sunspot drawings for each year in list form. Figure 4 [Figure 4: see original paper] shows the data display interface for 1991. Double-clicking any drawing in the query results displays the image and corresponding digital information, as shown in Figure 5 [Figure 5: see original paper], with mouse interaction enabling local image magnification in the right display box.

3.2 Sunspot Information Query

The Sunspot Information Query module includes time range queries, sunspot group number queries, and comprehensive queries based on date, sunspot group number, longitude, latitude, total area of individual sunspot group, area of the largest sunspot, and other parameters. After processing by the Sunspot Information Query Controller on the server, the system returns sunspot information matching the filter criteria. As an example, using comprehensive query with the date range from January 1, 1990, to January 1, 2000, and total area of individual sunspot group between 0.5 and 1.5 yields the results shown in Figure 6 [Figure 6: see original paper].

3.3 Sunspot Information Statistics

The Sunspot Information Statistics module includes: sunspot group area variation curves (including total area of individual sunspot groups and area of the largest sunspots), sunspot group radius variation curves, total sunspot group area variations, and sunspot number variations. The system can process annual observational data as well as statistical analyses spanning multiple years or even a century. Using the 1991 sunspot data as an example, the system-generated total area variation is shown in Figure 7 [Figure 7: see original paper], and the sunspot number statistics are displayed in Figure 8 [Figure 8: see original paper].

3.4 Long-Period Sunspot Information Statistical Analysis

Currently, the system has imported hand-drawn sunspot information from 1993 to 2015. Using the system to statistically analyze this data, researchers can clearly and intuitively observe annual variations in sunspot-related physical parameters. Statistical analysis of sunspot-related attribute values from 1993 to

2015 produced the results shown in Table 3 .

Through data visualization, the total area of individual sunspot groups, area of the largest sunspots, and counts from Table 3 are plotted by year, as shown in Figure 9 [Figure 9: see original paper]. The statistical analysis of sunspot area and sunspot (group) counts in Figure 9 exhibits clear wave-like peaks and troughs. The number of sunspots and sunspot groups correlates with the variation trend of sunspot area, where the magnitude of sunspot area and the number of sunspots directly characterize the intensity of solar activity. This prominent feature aligns with solar cycle variation patterns, which is significant for solar activity research and demonstrates that the data stored in the system is scientifically sound and reliable.

The Yunnan Observatories Hand-drawn Sunspot Data System has completed construction of the sunspot drawing database, sunspot information query, and sunspot information statistical analysis functions. The system enables researchers to conveniently query sunspot information under different conditions and conduct visualized studies of sunspot-related activities. The system functionality can be further expanded to provide convenient services for solar physics researchers and is significant for advancing sunspot activity research. Future work will focus on extending system functions to include sunspot motion trajectory mapping and three-dimensional sunspot image display, enabling more convenient, intuitive, and clear application to long-period sunspot activity research.

Acknowledgments

We thank Yunnan Observatories of the Chinese Academy of Sciences for providing the hand-drawn sunspot image data, and we are grateful to the relevant staff for their suggestions and support.

References

[1] Li Kejun, Su Tongwei, Liang Hongfei. Periodicity of sunspot activity observed by modern sunspots [J]. Chinese Science Bulletin, 2004, 49(24): 2511-2516.

[2] Baranyi T, Gyori L, Ludmány A, et al. Comparison of sunspot area data bases[J]. Monthly Notices of the Royal Astronomical Society, 2001, 323(1): 223-230.

[3] Győri L. Study of differences between sunspot and white light facular area data determined from SDO/HMI and SOHO/MDI observations[J]. Solar Physics, 2012, 280(2): 365-378.

[4] Balmaceda L A, Solanki S K, Krivova N A, et al. A homogeneous database of sunspot areas covering more than 130 years[J]. Journal of Geophysical Research Atmospheres, 2009, 114(A7): 431-433.

[5] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks[J]. Journal of Machine Learning Research, 2011, 15: 315-323.

[6] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[7] Yu D, Deng L. Deep learning and its applications to signal and information processing [Exploratory DSP][J]. IEEE Signal Processing Magazine, 2011, 28(1): 145-154.

[8] Niu X X, Suen C Y. A novel hybrid CNN-SVM classifier for recognizing handwritten digits[J]. Pattern Recognition, 2012, 45(4): 1318-1325.

[9] Zheng S, Zeng X Y, Lin G H, et al. Sunspot drawings handwritten character recognition method based on deep learning[J]. New Astronomy, 2016, 45: 54-59.

Note: Figure translations are in progress. See original paper for figures.