

CONSTRUCTION OF A CENTURY SOLAR CHROMOSPHERE DATA SET FOR SOLAR ACTIVITY RELATED RESEARCH

Ganghua Lin

*Key Laboratory of Solar Activity, National Astronomical
Observatories, Chinese Academy of Sciences,
Beijing, China, lgh@nao.cas.cn*

Xiao Fan Wang

*Key Laboratory of Solar Activity, National Astronomical
Observatories, Chinese Academy of Sciences,
Beijing, China*

Xiao Yang

*Key Laboratory of Solar Activity, National Astronomical
Observatories, Chinese Academy of Sciences,
Beijing, China*

Suo Liu

*Key Laboratory of Solar Activity, National Astronomical
Observatories, Chinese Academy of Sciences,
Beijing, China*

Mei Zhang

*Key Laboratory of Solar Activity, National Astronomical
Observatories, Chinese Academy of Sciences,
Beijing, China*

Haimin Wang

*Space Weather Research Laboratory, New Jersey Institute
of Technology,
Newark, USA*

Chang Liu

*Space Weather Research Laboratory, New Jersey Institute
of Technology,
Newark, USA*

Yan Xu

*Space Weather Research Laboratory, New Jersey Institute
of Technology,
Newark, USA*

A.G. Tlatov

*Kislovodsk Mountain Astronomical Station of the Pulkovo
observatory, Russian Academy of Sciences,
Kislovodsk, Russia*

M.L. Demidov

*Institute of Solar Terrestrial Physics SB RAS,
Irkutsk, Russia*

A.V. Borovik

*Institute of Solar Terrestrial Physics SB RAS,
Irkutsk, Russia Federation*

A.A. Golovko

*Institute of Solar Terrestrial Physics SB RAS,
Irkutsk, Russia*

Abstract. This article introduces our ongoing project “Construction of a Century Solar Chromosphere Data Set for Solar Activity Related Research”. Solar activities are the major sources of space weather that affects human lives. Some of the serious space weather consequences, for instance, include interruption of space communication and navigation, compromising the safety of astronauts and satellites, and damaging power grids. Therefore, the solar activity research has both scientific and social impacts. The major database is built up from digitized and standardized film data obtained by several observatories around the world and covers a timespan more than 100 years. After careful calibration, we will develop feature extraction and data mining tools

and provide them together with the comprehensive database for the astronomical community. Our final goal is to address several physical issues: filament behavior in solar cycles, abnormal behavior of solar cycle 24, large-scale solar eruptions, and sympathetic remote brightenings. Significant progresses are expected in data mining algorithms and software development, which will benefit the scientific analysis and eventually advance our understanding of solar cycles.

Keywords: solar cycle, $H\alpha$, filament, multi-parameter calibration, standardization, feature extraction, solar activity pattern.

BACKGROUND

The Sun is the only main sequence star with periodic activities that dominates the Sun-Earth environment and has impacts on human lives. Strong solar activities such as CMEs can affect spacecraft systems, disrupt communications, and even damage ground-based power systems. To understand and forecast the effects of solar activity on Earth’s environment remains one of the main research problems in solar physics. However, our understanding of the fundamental physics of solar activities is still poor. For example, the mechanism of formation of solar cycle variation is unclear so far. In the literature, the Maunder minimum is well-known, and the abnormally depressed solar activity between cycles 23 and 24 has re-

cently started to be discussed. However, their occurrences are big puzzles to solar physicists. On the other hand, the existing observational data are not well calibrated and organized to serve efficient and precise investigations.

In particular, long-term observations are needed to study variation of solar activities in multiple solar cycles. Such activities include large-scale eruptions, Moreton waves, sympathetic eruptions, remote brightening [Tang, Moore, 1982], coronal dimming, filament oscillations, and so on. Systematic studies of these large-scale events are essential to understand the topological magnetic structures and hence the eruptive events.

The Moreton wave is a disturbance propagated by a wave and generally accompanied by a flare. It was first

detected on H α filtergrams by Moreton, Ramsey [1960]. So far, since Moreton wave events are not commonly observed, it is necessary to proceed to a more systematic analysis by taking advantage of the large H α data sample all over the world.

The planned database is extremely useful in investigating large-scale helicity patterns, which are believed to have a close relationship with solar activities. For example, it can be used to study the possible sign reversal problem in the hemisphere helicity rule. (e.g., [Bao, Ai, Zhang, 2000; Hagino, Sakurai, 2002; Pevtsov et al, 2008]).

It is necessary to establish a complete filament catalog in multiple solar cycles. The statistics study of filament properties in many cycles is restricted by discontinuous observations, inconsistent calibration, and incomplete samples from different instruments. It is well known that filaments typically fall into three categories [Hansen R., Hansen S., 1975] according to their latitudes. We pay attention to large-scale filaments in middle and high latitudes as they are high contrast features and their properties are more closely connected with solar cycle problems [Brajša et al. 1990]. Pevtsov, Balasubramaniam, Rogers [2003] studied chirality of chromospheric filaments, using a limited data set of scanned images from the NSO/SP database. In this project, we further promote the research into solar helicity and chirality problems (e.g., [Martin, Bilimoria, Tracadas, 1994; Rust, Martin, 1994; Pevtsov, Canfield, Metcalf, 1995]), through a complete, unified, and calibrated digitized full-disk H α data set.

The database covering multiple solar cycles is being constructed in stages. In our project, by combining the large-scale structure of solar eruptions, the long-term characteristic variation of filament statistics, and the abnormal behavior of solar cycles, the century-long nine solar cycles' chromospheric data set around the world will be integrated. Upon accomplishment, it can provide strong support for further analysis of solar activities and can serve as the foundation for developing methods and references for deeply exploring the scientific values of astronomical observations. Our project can give theoretical support to the space weather forecast, which is becoming more and more urgent; and tools developed can be applied to the national space weather forecast system. Hence, this project has an important scientific value and social significance.

Meanwhile, information and feature extraction is the basis for the long-cycle data set building process. In the SDO era, observation data already have the features of big data, namely the characteristics of mega data (volume, variety, velocity, variability, veracity, value, and complexity). For example, browsing images one by one from a data set with a second-sample rate and trying to find the phenomena of interest are beyond the ability of a personal computer system. Therefore, big data characteristics have posed a new challenge to the data analysis. To analyze and process SDO data, the United States, the European Union, and other countries have developed many data mining tools for all kinds of solar features and solar event parameters. Until now, there are seven special workshops held to address the key subject of solar image

data and processing. The stability, accuracy, and efficiency of solar information extraction and corresponding tools are the main topics of these workshops. Automatic processing methods for massive amounts of data are also discussed in these workshops [Aschwanden, 2010].

Carried out for more than 100 years (at the Kodakikanal Observatory built in 1899, H α observation began in 1912), full-disk H α observations have provided a significant amount of data. However, there was no internationally accepted standard for calibration until Zharkova et al. [2002] published their H α data standardization process. The authors described the correction of non-uniformity of the disk shape and intensity, using data from the Meudon Observatory in France (a member of the Global High Resolution H α Network). Ermolli et al. [2009] studied Calcium plage areas over nine solar cycles, using full-disk Ca II K data from three observatories. The obvious discrepancy between the data sets demonstrates that a unified calibration standard for full-disk solar images is necessary but difficult to establish.

In terms of extraction of solar data, implementations have been developed in recent years based on different models such as Bayes classifier, wavelet transform, morphology [Cui Zhao et al., 2016], artificial neural network, support vector machine (SVM) [Qu M. et al., 2003], deep learning [Sheng Zheng et al., 2016], and so on. Sometimes, better results can be obtained by combining several methods.

In recent years, significant progress has been made in filament detection and statistics studies of filaments and their eruptions. Yuan et al. [2011] presented a method for filament recognition and feature extraction, using examples of 125 H α images, mainly from the Big Bear Solar Observatory (BBSO) and other three stations (the samples are not large: 10 years, one image per month; the filament skeleton process and calculation are fulfilled, but the chirality calculation are not). Hao, Fang, Chen [2013] have studied 13832 filaments from 3470 H α images obtained by the Mauna Loa Observatory during the solar cycle from 1998 to 2009. Hao [PhD Thesis, 2015] extends the analysis to cover three solar cycles based on BBSO H α data; many algorithms have been developed to calculate the interconnection between filament fragments, filament spine, tilt angle, chirality (not solid definition), evolution tracking, and migration statistics. Tlatov et al. [2016] published the results of filament tilt angles and distribution of filament butterfly diagrams for nine solar cycles. The data set of Tlatov et al. [2016] comprises historical synoptic charts obtained manually from digitized data. It should be noted that the manual method is still the most accurate way to draw such charts, but it is time consuming.

In summary, to serve astronomy research purposes, we plan to integrate as many existing long-duration observations from different observatories as possible. Combining and calibrating historical data make up an important direction of solar and astronomical data processing. "Intensity calibration, geometric distortion, resolution rescale, large scale inhomogeneity (include cloud problem), etc." can be expected to be the main difficulties in this project.

H α archive overview

Station	Country	Period	Approximate FOV	Band
Kodaikanal	India	1912–2006	Full disk	H α
Meudon	France	1919–2003	Full disk	H α
Sac Peak	US	1953–2003	Full disk	H α
Kislovodsk & Baikal	Russia (USSR)	1958– now	Full disk	H α
BBSO	US	1971–1995 1973–1995 1969–1995	Full disk 200"×150" 400"×300"	H α (with off-band), magnetogram
Nanjing	China	2011– now	Full disk	H α
YNAO	China	1981– now	Full disk	H α
HSOS	China	2000– now	Full disk	H α (with off-band), magnetogram
GHN	International Cooperation (including China)	2000– now	Full disk	H α

In addition to the new data archive, algorithms and tools are being developed for information extraction, especially for automatic recognition of relevant astronomical events.

At present, due to the close cooperation among BBSO, NSO/SP, Kodaikanal Observatory, and GHN, century-long H α data sets have been collected and combined successfully. The former Soviet Union H α data archives, methods and tools will be included in our future cooperation.

2. MAIN RESEARCH CONTENTS

The technical component of our project involves integrating multiple databases globally (primarily full-disk H α) and developing methods and tools for automatic feature detections. The scientific products are based on the technical component and address several important issues in solar physics such as the global scale solar eruption, remote brightening, Moreton wave, multi-cycles filament property statistics, and abnormal behavior of solar cycles 23 and 24. These contents can be distributed into three stages: to establish the comprehensive database and preprocess full-disk images; to develop methods and tools for filament feature recognition; to discover recurring patterns of solar activities. Below are the detailed descriptions of each task:

Comprehensive database

Firstly, time stamps are recovered. Before CCD cameras became popular in astronomical observations, the widely used recording medium was film or plate. The observing time was photographed as digital clocks on images. As shown in Table, the data volume is huge, spanning multiple solar cycles. Therefore, manually recognizing the time stamps is an extremely time-consuming task. An automatic method for recovering the observing time has been developed and started to be applied to full-disk images. Secondly, multi-parameter calibration, normalization and FITS standardization are performed. Historical data are collected from different facilities, and different digitizers are used. Therefore, the pixel resolution and intensity need to be unified in

the recreated data sets. Another task is to correct the disk distortion, introduced by the digitizers and film deformation, by a re-mapping method based on empirical models. There is another minor issue of Carrington coordinates. Sometimes there is a gap, about few minutes, between two consecutive Carrington rotations. This is a systematic uncertainty that could be introduced in such a long timespan of multiple solar cycles. After these calibrations, all of the images are converted to the standard FITS format and are made available to public. We have built up a 200 TB data server to store the data sets. A server for data processing will be created for scientific data calibration and processing. For instance, the processing server will be able to carry out parallel data processing, which is important for pattern recognition.

To develop methods and tools for filament detection

We plan to develop a comprehensive information framework that mainly focuses on various parameters of filaments. This will help to detect filaments and other solar activities automatically.

Statistical investigation of solar activities

By implementing the new method and tools in the comprehensive database, we plan to address the scientific issues that are closely related to temporal variations of solar activities; for instance, statistical properties of filaments in super-long solar cycles, abnormal behavior in cycles 23 and 24, halo-CMEs, Moreton wave, and sympathetic eruptions.

3. RESEARCH OBJECTIVES

Our major research objectives are as follows:

- 1) Building up a unique database covering nearly 100 years and setting up a standard for calibrating international H α data;
- 2) Developing a sophisticated algorithm to retrieve multiple physical parameters;
- 3) Understanding the abnormal behavior between solar cycles 23 and 24; figuring out more principles about the long-term distribution of filaments; studying large-scale structures during solar eruptions.

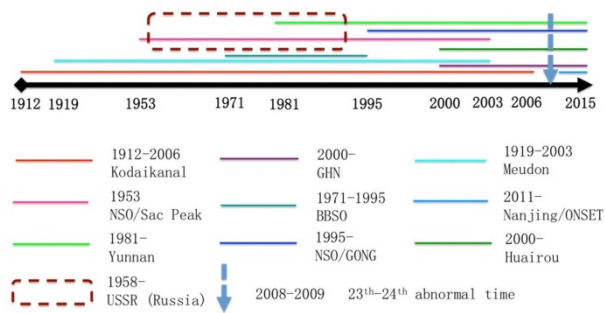


Figure. Ha observing time coverage

4. ABOUT RUSSIA H α DATA

Russia, the largest country, covers nine time zones. Without considering the longitudes covered by the sea, more than half of the terrestrial longitudes are within Russian territory. Figure shows the global distribution status of H α observing time coverage for one century.

The historical database from Russia is mutually complementary to those from other stations and therefore plays an important role. Hence, it is indispensable to our project (the Russian partners were not involved in the Chinese NSFC project of No. U1531247 because the former Soviet Union historical H α data conditions were unclear at the time we composed the project).

The work is jointly supported by the Chinese Academy of Sciences and Natural Science Funds of China (No. U1531247), Ministry of Science and Technology of China (No. 2014FY120300, No. 2012FY120500). Many of scientific proposals in this project are originated from Prof. Haimin Wang's instruction and suggestion. Related works were partially supported by Natural Science Funds of China (No. 11103041, No. U1331113, No. 11573042) and NSF (No. 1620875) and NASA (No. NNX16AD67G). We are grateful to Prof. Pevtsov (Russia, NSO/USA), Prof. W.PoTzi (KOSER/Austria), and Prof. Dipankar (Kodaikanal Observatory/India) for providing their data during the data gathering.

REFERENCES

Aschwanden M.J. Image Processing Techniques and Feature Recognition in Solar Physics. *Solar Phys.* 2010, vol. 262, pp. 235–275.

Bao S.D., Ai G.X., Zhang H.Q. The Hemispheric Sign Rule of Current Helicity During the Rising Phase of Cycle 23. *J. Astrophys. Astron.* 2000, vol. 21, pp. 303–306.

Brajsa R, Vršnak B, Rundjak V., Schroll A. Polar Crown Filaments and Solar Differential Rotation at High Latitudes. *IAU Colloquim.* 1990, pp. 117–293.

Cui Zhao, GangHua Lin, YuanYong Deng, Xiao Yang. Automatic Recognition of Sunspots in HSOS Full-Disk Solar Images. *Publications of the Astronomical Society of Australia.* 2016, vol. 33. DOI: 10.1017/pasa.2016.17.

Ermolli I., Solanki S.K., Tlatov A.G., Krivova N.A., Ulrich R.K., Singh J. Comparison among CaII K spectroheliogram time series with an application to solar activity studies. *Astrophys. J.* 2009, vol. 698, pp. 1000–1009.

Hansen R., Hansen S. Global distribution of filaments during solar cycle No. 20. *Sol. Phys.* 1975, vol. 44, pp. 225–230.

Hagino M., Sakurai T. Hemispheric helicity asymmetry in active regions for solar cycle 21–23. *Proc. COSPAR Colloquia Ser.* 2002, vol. 147.

Hao Q., Fang C., Chen P.F. Developing an advanced automated method for solar filament recognition and its scientific application to a solar cycle of MLSO H α data *Solar Phys.* 2013, vol. 286, pp. 385–404.

Moreton G.E., Ramsey H.E. Recent observations of dynamical phenomena associated with solar flares. *Publications of the Astronomical Society of the Pacific.* 1960, vol. 72, pp. 357–358.

Martin S.F., Bilimoria R., Tracadas P.W. Magnetic field configurations basic to filament channels and filaments. *Solar surface magnetism. NATO Advanced Science Institutes (ASI) Ser. C.: Mathematical and Physical Sciences, Proc. NATO Advanced Research Workshop.* 1994, vol. 303.

Pevtsov A.A., Canfield R.C., Metcalf T.R. Latitudinal variation of helicity of photospheric magnetic fields. *Astrophys. J. Lett.* 1995, vol. 440, pp. L109–112.

Pevtsov A.A., Balasubramaniam K.S., Rogers J.W. Chirality of chromospheric filaments. *Astrophys. J.* 2003, vol. 595, pp. 500–505.

Pevtsov A.A., Canfield R.C., Sakurai T., Hagino M. On the solar cycle variation of the hemispheric helicity rule. *Astrophys. J.* 2008, vol. 677, no. 1, pp. 719–722.

Qu M., Shih F.Y., Jing J., Wang H. Automatic solar flare detection using MLP, RBF, and SVM. *Solar Phys.* 2003, vol. 217, pp. 157–172. DOI: 10.1007/s11207-013-0285-9.

Rust D.M., Martin S.F. A correlation between sunspot whirls and filament type. *Astronomical Society of the Pacific Conference Ser.* 1994, vol. 68, pp. 337.

Sheng Zheng, Xiangyun Zeng, Ganghua Lin, Cui Zhao, Yongli Feng, Jinping Tao, Daoyuan Zhu, Li Xiong. Sunspot drawing handwritten character recognition method based on deep learning. *New Astronomy.* 2016, vol. 45, pp. 54–59.

Tang F., Moore R.L. Remote flare brightenings and type III reverse slope bursts. *Solar Phys.* 1982, vol. 77, pp. 263–276.

Yuan Y. Shih F.Y., Jing J., Wang H., Chae J. Automatic Solar Filament Segmentation and Characterization. *Solar Phys.* 2011, vol. 272, pp. 101. DOI: 10.1007/s11207-011-9798-2.

Zharkova V.V. et al. A full disk image standardization of the synoptic solar observation at the Meudon. *ESA SP-506.* 2002, vol. 2, pp. 975–978.

How to cite this article

Ganghua Lin, XiaoFan Wang, Xiao Yang, Suo Liu, Mei Zhang, Haimin Wang, Chang Liu, Yan Xu, Tlatov A.G., Demidov M.L., Borovik A., Golovko A. Construction of a century solar chromosphere data set for solar activity related research. *Solar-Terrestrial Physics.* 2017. Vol. 3. Iss. 2. P. 5–8.