Library and Information Services in Astronomy VII: Open Science at the Frontiers of Librarianship
ASP Conference Series, Vol. 492
A. Holl, S. Lesteven, D. Dietrich, and A. Gasperini, eds.

Retrospective Conversion of Solar Data Printed in "Synoptic Maps of the Solar Chromosphere": A Scientific and Librarianship Project

Amélia Laurenceau,[1] Jean Aboudarham,[2] and Christian Renié[2]

[1]*Observatoire de Paris, UMS 2201 CNRS, PSL Research University, 61 avenue de l'Observatoire, 75014 Paris, France*
[2]*LESIA, Observatoire de Paris, UMR 8109 CNRS, UPMC, Université Paris Diderot, PSL Research University, 5 place Janssen, 92195 Meudon, France*

Abstract. Between 1928 and 2003, the Observatoire de Paris published solar activity maps and their corresponding data tables, first in the Annals of the Meudon Observatory, then in the Synoptic Maps of the Solar Chromosphere. These maps represent the main solar structures in a single view and spread out on a complete Carrington rotation as well as tables of associated data, containing various information on these structures such as positions, length, morphological characteristics, and behavior. Since 2003, these maps and data tables have not been released in print, as they are only published on the online BASS2000 database, the solar database maintained by LESIA (Laboratory for space studies and astrophysical instruments). In order to make the first 80 years of observations which were available only in paper accessible and usable, the LESIA and the Library of the Observatory have started a project to digitize the publications, enter the data with the assistance of a specialized company, and then migrate the files obtained in BASS2000 and in the Heliophysics Features Catalog created in the framework of the European project HELIO.

1. Introduction

We will first introduce the historical background of the project to make the solar activity maps and their corresponding data tables available and usable online. Then, we will present the different stages of the project and the methodology used to show how the scientific and technical skills of an astronomer, a librarian, and an IT specialist helped to carry out the project. Finally, we will discuss the scientific opportunities offered by this free provision of the oldest and longest series of solar observations.

2. Meudon's spectroheliograph

Henri Deslandres (1854–1948) was appointed an astronomer in Paris in 1890. In 1891, he invented, at the same time as American solar astronomer George Ellery Hale (1866–1938), the spectroheliograph. This is an instrument capable of capturing photographic images of the sun at a single wavelength of the light spectrum. It works with a spectrograph, a coelostat, and a siderostat. Deslandres's first spectroheliograph was used in 1893 in Paris (see Figure 1), then in Meudon, where he was appointed in 1897.

When Jules Janssen (1824–1907) died in 1907, Deslandres became director of the Meudon Observatory. The same year, the first congress of the International Union for cooperation in solar research took place. In 1909, a new, large, 14-meter long spectroheliograph (see Figure 1), conceived by Deslandres and his assistant Lucien d'Azambuja (1884–1970) was introduced: there began the Meudon spectroheliogram collection, which is the oldest in the world. Observations were conducted regularly, with a four-year interruption during the First World War. March 15th, 1919 marked the beginning of a series of daily observations of the hydrogen Hα and CaK lines (K3 centre of the line, and K1v violet line wing), to be supplemented by images from Mount Wilson (USA), Kodaikanal (India), and Coimbra (Portugal) Observatories.

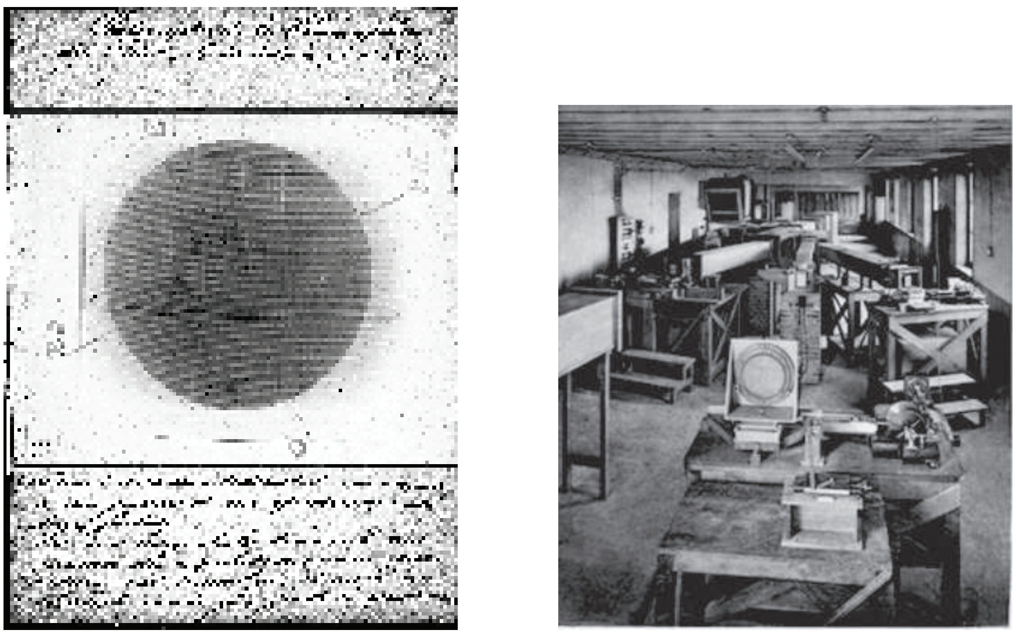

Figure 1. Left panel: The first spectroheliogram, by H. Deslandres (Oct. 24th 1893) - Right panel: Large spectroheliograph with a 14 meter-long chamber (left) Meudon Observatory, 1909 Henri Deslandres

The same instrument is used today to guarantee homogeneity in the observations. It was refurbished in 1989 by G. Olivieri who changed some optical and mechanical pieces, and computerised its use in 2003 by adding a CCD camera. The instrument's principle characteristics have been preserved so as not to alter the consistency of the collection.

3. Synoptic maps of the solar chromosphere

Spectroheliograms allow the main solar structures visible in the chromosphere, particularly filaments, prominences, chromospheric faculae and even sunspots in the photosphere, to be recorded. The structures shown on each spectroheliogram are measured and the corresponding data (e.g., morphological, position, and behaviour information) are compiled in the data tables. From these data, one can then draw synoptic maps representing the sun projection for a complete solar rotation, including the main solar structures in their most remarkable position in the course of the rotation. The maps give us a snapshot of the maximum activity happening during one solar rotation. An example data table and synoptic map is given on Figure 2.

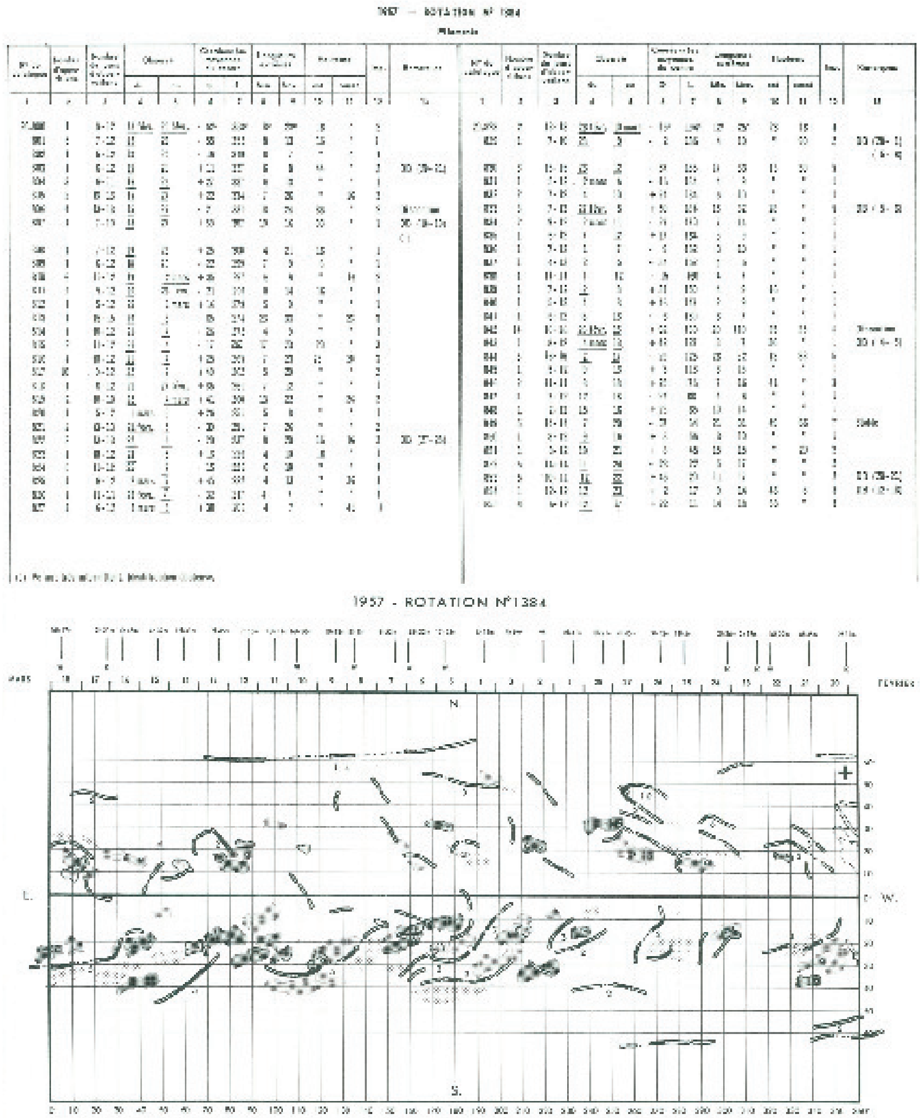

Figure 2. Upper panel: data table for filaments. Lower panel: the corresponding synoptic map, including chromospheric faculae and sunspots, in addition to filaments

The type of data that has been compiled and the method used to make the maps have evolved with time. Deslandres et D'Azambuja (1913) first presented a graphical representation that allowed them to follow the evolution of the solar filaments before the Académie des Sciences. According the minutes of the weekly meetings of the Académie des Sciences, D'Azambuja (1921) introduced a map showing all filaments that can be seen during a solar rotation: this map was the ancestor of synoptic maps. The following year, the UAI general assembly adopted d'Azambuja's model of synoptic map representation. From 1923, Marguerite Roumens, who was in charge of measuring the spectrums and who became Mrs. d'Azambuja in 1935, helped him. In 1928, the first maps for the 1919–1920 period were published in the Annales de l'Observatoire. They included the identification and the measurements of 575 filaments, thus inaugurating the beginning of a new series of publications for the following periods indicated in Table 1.

Table 1. Content of publications

Period of time	Rotation	Publication
1919–1930	876–1033	Annales de l'Observatoire de Paris. Section de Meudon ISSN : 1249-786X - Vol. 6 n 1–6 ; 1926–1933
1931–1989	1034–1823	Cartes synoptiques de la chromosphère solaire et catalogue des filaments et des centres d'activité - ISSN : 0373-7713 - Vol.1–7 ; 1934–1991
1990–1991 ; 1998–2001	1824–1850 ; 1931–1984	Cartes synoptiques de l'activité solaire ISSN : 1290-0230 - Vol. 8–10 ; 1998–2003
1990–2003	1824–2008	BASS2000

At first, maps were entirely drawn by hand. The method evolved in 1950 when Roger Servajean invented an anamorphoseur (shown on Figure 3), allowing the conversion of spherical coordinates into cylindrical or surface coordinates. The instrument made the conversions easier to calculate and improved the precision of the measurements.

In 1990, a semi-automated method was implemented by Zadig Mouradian Mouradian (1998): the images were scanned and the solar structures were identified using software. Human decisions are still necessary, so the process is not fully computerised. The maps and data for the 1990–2003 period were published in BASS2000.[1] This method was used until 2003, when Germaine Zlicaric — who was in charge of entering the data, analysing them, and sketching the maps — retired.

Since 2003, automatic structure-recognition computer programmes have been developed, such as the one for filaments by Fuller et al. Fuller et al. (2005). Because of these programmes, it has been possible to extract data from photographs and to make them available in the Heliophysics Feature Catalogue and in BASS2000. These made the Meudon observatory famous for its solar observations, as BASS2000 became the World Data Center for Solar Activity in 2012. But synoptic maps are not drawn anymore and work is being done to fully automate their creation.

4. Preserving and rendering 80 years of solar observations to be accessible to academics

As we have previously seen, there are synoptic maps and associated data in paper form from 1919 to 2001. This represents more than 80 years of data, meaning that there are about 853,000 data points for 7 full solar cycles which only exist in printed form and

[1] http://bass2000.obspm.fr/

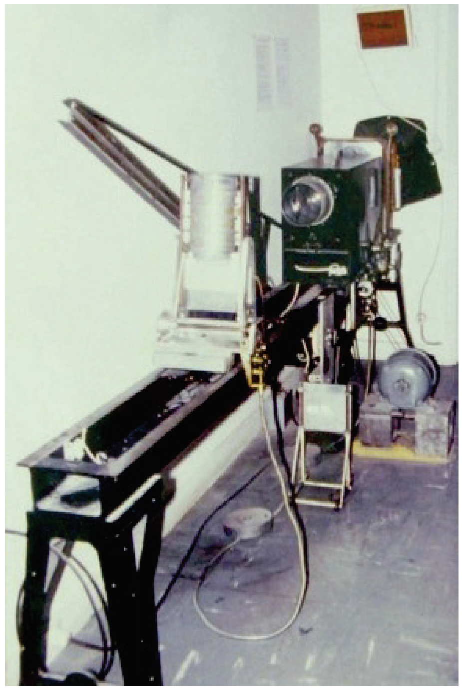

Figure 3. “Anamorphoseur” aimed at facilitating the making of synoptic maps of the solar chromosphere from spectroheliogram images Mechanics Workshop, Meudon Observatory, 1949 M. Brebion

which consequently cannot easily be utilized. The LESIA and the Observatoire de Paris Library teamed up in a rescue project to make these data available. The Observatoire de Paris Scientific Board funded this project.

The project involved getting a specialized company to input the data from the PDF files obtained by digitizing printed publications. Two different people enter the data independently, and if there is a difference in what each person entered, a third person checks and corrects the data. The second step of the project consists in migrating these files into the solar observation databases (BASS2000 and Heliophysics Feature Catalogue) so the scientific community can access them freely.

4.1. Entering the data

4.1.1. Identifying, understanding, describing and locating the data

The first step of the project consisted in identifying the different table models and, for each of them, understanding the nature of the data. A table model was created for each set of tables that present the same intellectual structure, i.e. the data is structured the same no matter what the table looks like in its printed form. Then, a migration script can be applied to all files containing the data entered for the same table model. To identify 28 different table models, 4.887 pages had to be read. These tables have then been precisely described, so the company providing the data input service could recognise them.

The table models are unequally represented: there is a table model which only appears once and on one page, while other models may appear on a thousand pages and represent significantly more data. For each model, the company had to be given the location of the tables in the booklets. This task was a difficult one, because on most publications there is no pagination available and tables are often printed alternately with synoptic maps.

4.1.2. Data entry instructions

To guarantee the highest quality possible results, the data entered had to be as close to the original data as possible. However, character strings that could not be entered as they were, such as the ones listed below, had to be identified.

- All characters which did not exist in the ASCII table,
- characters indicating an absence of data,
- characters indicating a repetition of the previous data,
- and all other particular cases.

4.1.3. Defining technical specifications

A specifications document had to be drafted for the data input company. These technical specifications included:

- The description of the different data models and their location;
- data-entering instructions;
- a naming plan for the delivered files; and
- a way to trace data, so that for all data entered, the originating page and pdf file was known.

A call for tenders was sent out to companies that specialised in information handling in January 2014. Six companies applied. At the end, INOVCOM was chosen for its excellent quality rate (the company is committed to providing 99.99% accuracy) at a reasonable price.

4.1.4. Quality control

There is still a need for quality control measures, even with a 99.99% data accuracy rate. At the same time as the technical instructions were given, data-control programmes were elaborated. These were based on examining the internal consistency of them data. Examples include:

- The columns containing degree measures must always have a number comprised between 0 and 360;
- some columns are the sum of the two previous columns; and
- the values in all columns showing filament numbers follow each other (n+1).

4.1.5. Testing

The first input test for each table model was done at the end of May 2014 to make sure that the company knew the different table models and all of the associated instructions. The test analysis is in progress. The test showed that there were mistakes in the print documents: we also will have to correct the original printed data. If the test is validated, 5% of the data will be entered. If this first part is validated, production should take place from the end of July through October.

4.2. Moving the data into solar databases

The next step will be the creation of a database to gather and review all the data entered. This will be done in two steps. In the first step, the main data (e.g., observation date, position, dimensions, etc.) will be incorporated into pre-existing tables of the Heliophysics Feature Catalogue so as to augment its content up to eight 11-year solar cycles instead of the only one currently available. In the second step, specific tables will have to be created for the oldest data because the data structure does not match with the Heliophysics feature Catalogue Standards.

5. Complementary skills: a scientist, a computer scientist, and a librarian

For the project to be successful, several types of skills were required:

- Scientific skills, to identify the nature and interest of the data and control their consistency;
- information science skills; to create data structures;
- computer skills; for programming and data migration; and
- project management skills.

The project is carried out by a solar physicist (Jean Aboudarham), a computer scientist (Christian Renié), and a librarian (Amélia Laurenceau), under the scientific direction of Jean Aboudarham. Jean Aboudarham conducted a scientific analysis of the data, Christian Renié was in charge of the data migration, and Amélia Laurenceau was in charge of the technical instructions and the project management for the data input. In terms of information science, the project is, in many aspects, similar to a retrospective conversion. It is indeed about converting structured data from print to an electronic format.

6. Which scientific perspectives?

The main solar structures have been scientifically studied for about a century, but the analysis of their behaviour was limited to very short periods of time: the best case was an eleven-year solar cycle, with the sun's positions recorded on graphs or maps. In most cases, detailed studies were limited to one solar rotation (28 days). The availability of these data will allow for scientific studies that have never been possible until now. Some examples, include:

- a very long-term study of these structures' behavioural tendencies;

- a study of the possible correlations between the behaviours of different structures;
- a study of the tendencies related to the different phases (crescent, waning) of the solar cycle over the course of 8 cycles, which allows us to start building up a statistical record;
- a statistical study of solar activity over several cycles; and
- a first-time study of the 80-year solar cycle about which we know very little from numerous solar structures;

Finally, it is likely that the discoveries arising from the points mentioned above will lead to new questions and new research paths in the years to come.

7. New projects

Many observation data still only exist in printed form, whether manuscripts, typescripts (observation records), or printed publications. Digitising these data would be useful for preservation purposes. Similar projects to those that still have a scientific interest could be carried out so as to make their data available in scientific databases.

Our team has already started on a new project of retroactive conversion of the data published in the Quarterly Bulletin of Solar Activity (1939–). The availability of these data would provide a listing of solar flares. From this listing, it is possible to study the correlations between the structure's behaviours and the flares, and even determine the warning signs of solar flares, so as to more efficiently protect the technological devices sent out into space.

References

Aboudarham, J., Scholl, I., Fuller, N., Fouesneau, M., Galametz, M., Gonon, F., Maire, A. & Leroy, Y. 2007, in ASP Conf. Ser. Vol. 368, (San Francisco: ASP), 505

D'Azambuja, L. 1921, Comptes rendus hebdomadaires des séances de l'Académie des sciences publiés... par MM. les secrétaires perpétuels, 173, 1450

Deslandres, H. & D'Azambuja, L. 1913, Comptes rendus hebdomadaires des séances de l'Académie des sciences publiés... par MM. les secrétaires perpétuels, 157, 413

]dollfus05 Dollfus, A. 2005, L'Astronomie, 119, 150

Fuller, N., Aboudarham, J. & Bentley, R.D. 2005, Solar Physics, 227, 61

Martres, M.J. 1998, in ASP Conf. Ser. Vol. 150, (San Francisco: ASP), 3

Mouradian, Z. 1998, in ASP Conf. Ser. Vol. 140, (San Francisco: ASP), 197