Mem. S.A.It. Vol. 76, 862

Memorie della

The solar photograph archive of the Mount Wilson Observatory

A resource for a century of digital data

S. Lefebvre, R. K. Ulrich, L. S. Webster, F. Varadi, J. Javaraiah, L. Bertello, L. Werden, J. E. Boyden and P. Gilman

University of California at Los Angeles, Department of Physics and Astronomy, Physics and Astronomy Building, 430 Portola Plaza, Box 951547, Los Angeles, CA 90095-1547, USA e-mail: ulrich@astro.ucla.edu

Abstract. The solar telescopes and spectroheliographs of the Mount Wilson Observatory were among the earliest modern facilities for the study of the solar surface. The photographic collection of the solar program at Mt. Wilson begins in 1894 and continues to the present day. A program to digitize and distribute the images in this collection was begun at UCLA in 2003 and is now making available the first of the catalogued and catagorized images from the CaK sequence. Most of the instrumentation with which the images were obtained is still available although in a disassembled form. Original log books have been digitized and associated with the images so that a maximum of scientific return can be obtained from the data base. The present range of images available from www.astro.ucla.edu/ ~ulrich extends from late 1915 to 1952. Each image has been digitized with 12-bit precision and represented in a 16-bit format. These images are each 13 Mbytes in size and larger than will be the final product images since not all image defects have been mitigated at this time. The radii and centers of the solar images have been determined and are included in the available data files. Optical vignetting by the system introduces an intensity gradient of known magnitude that can be used to help characterize the photograph plates. The roll angle of the images has yet to be determined.

Key words. Sun: historical data – Sun: variability – Sun: rotation – Sun: magnetism

1. Introduction

The Mt. Wilson Solar Photographic Archive DIgitization Project (Mt. Wilson SPADIP) makes available to the scientific community in digital form solar images in the archives of the Carnegie Observatories. These images date from 1894 and include many which can be judged to be of superb quality by modern standards. These images will permit a variety of retrospective analysis of the state of solar magnetism and provide a temporal baseline of about 100 years for many solar properties. The 20th century was a period of increased anthropogenic production of greenhouse gases which might contribute to global climate change. The Sun could also be a factor in global climate

Send offprint requests to: R. Ulrich

change and the data provide by this digitization will allow the scientific community to freely examine some properties of solar magnetism over the 20th century. The objectives of the project are:

1. to preserve the original plate material from time degradation;
2. to provide a web-based digital archive with incorporated log book parameters;
3. and to calibrate the images as much as possible.

This digitized archive should permit the study of a wide range of topics including many that we cannot anticipate. The distribution of the images is unrestricted by means of a website and enable community utilization of this historical database. In addition, we have identified three specific problems that can be addressed with this database. We will carry out studies of the following questions:

1. How is the strength and distribution of the Sun's weak magnetic field related to the sunspot number?
2. What is the temporal history of the Sun's surface differential rotation?
3. What is the history of the strength and distribution sunspot darkening and facular brightening?

2. Type of data and management

2.1. History of the archive

Mt. Wilson solar Observatory was founded in 1904 by George Ellery Hale under the auspices of the Carnegie Institution of Washington. In that year, Hale brought the Snow Solar Telescope from Yerkes Observatory in southern Wisconsin to the sunnier and steadier skies of Mt. Wilson to continue his studies of the Sun. The 60-foot tower was the first tower configuration on Mt. Wilson. Hale used this telescope to identify magnetic fields in sunspots for the firs time.The program of Calcium spectroheliograms began prior to Hale's selection of Mt. Wilson for the observatory. The earliest plates were obtained at the Kenwood Observatory during a period between 1891 and 1895. The earliest plates are of variable quality. Beginning in 1905, several series of calcium spetroheliograms, whose A-series, were obtaines at the 60-foot tower telescope and at the Snow Telecope. And in 1905, broad-band images called white light directs are included. Digitized forms of these images are included in the digital Mt. Wilson solar photographic archive.

2.2. The plates and scans

The archive contains over 150,000 images of the Sun which were acquired over a time span in excess of 100 years. The glass and acetate negatives are stored and maintained at the Pasadena California offices of the Observatories of the Carnegie Institution of Washington. The images include over 43,000 broad-band images called White Light Directs, over 35,000 CaK spectroheliograms and over 74,000 Hα spectroheliograms. Figure 1 shows the coverage of the Calcium spectroheliograms from 1915 to 1952. Before 1940, there is an average of 400 images per year, and 700 after 1940.

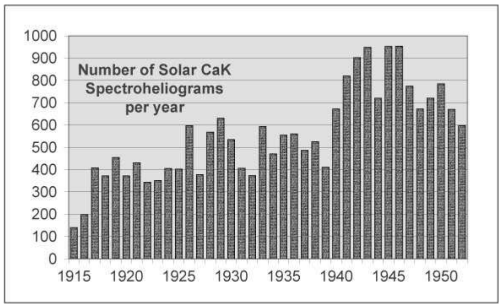

Fig. 1. Number of CaK spectroheliograms per year.

The plates are fragile and must be handled with care in order to preserve their quality. Prior envelopes, where plates were stored, have glue seams and higher acid content, the wooden cases outgas harmful chemicals. Now, plates are in archival 4-flap housing in inert boxes without glue. The photographic plates have been scanned in groups of three with each plate having up to 4 images. Most early plates have two pairs with each pair being exposed simultaneously. One image of the pair is at the center of the CaK line and includes the light of CaK2V, K3 and K2R. The second image of the pair is in the light of the nearby part of the K1 line and does not usually include other spectral features. We do not use the data from the second images although they are included in the general scans. The scans are being carried out using an Eskographics F14 scanner. The solar images have radii of about 1000 pixels and the digitization provides about 12 bits of significant data. The scanner output images are in a 16-bit TIF format with the intensity values being stored as positive integers. We refer to these initial scans as fulldeck scans since all images are scanned simultaneously using the full deck of the scanner.

2.3. Log book information and scan identification

Data from the original log books has been entered manually and examined for validity. These digitized log files are downloaded to an archive database. The solar images on the full deck scans are then tagged with corresponding digital log book entries by visually examining each image on the scan for written marking giving the date, time and observation ID number. This tagging step links the time and date of observation to the image position on the full deck scan. Using the tag information, an extraction is done with software written specifically for this purpose. Plate geometry and solar image geometry are essential in this step. Each extracted image is stored as a separate 16-bit TIF file and the information from the database is recorded as the figure caption of the image in a format that can be used during subsequent processing steps. This initial extraction provides the first indication of the location of the solar image since the extraction software finds a circular structure in the expected position and then picks a 2601 by 2601 pixel array with this circle of roughly 1000 pixels radius at its center.

2.4. The website

A web site has been developped and is always updated. The images are freely available through http://www.astro.ucla.edu/~ulrich/MW_SPADP/. Raw images as well as images subjected to partial processing are included. On this website, browseable images with thumbnails are presently available from 1915 to 1952 and permit easily to the user to choose the image of interest for him. An example of thumbnails is presented on Figure 2.

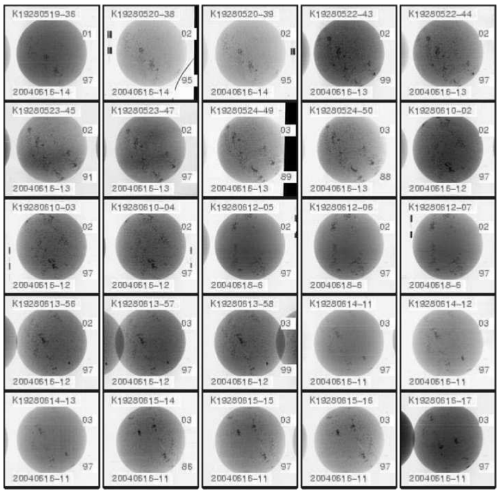

Fig. 2. Example of thumbnail issued from the SPADIP website.

Reduced-size science quality FITS files with image center as well as vertical and horizontal radii determined are also available for the same years. Summary listings of the radius solution results are found in each year's directory. Fits file description with a brief definition of the algorithms for finding the radii and image centers is provided. A listing of the quantities provided in the summary files in each yearly directory is accessible. The precised algorithms used are in FORTRAN and is also on the website.

Figure 3 shows an example of Calcium spectroheliogram taken on June 25th 1928. Notice the good quality of this negative image:

a lot of plages anf faculae can be seen even at higher latitude.

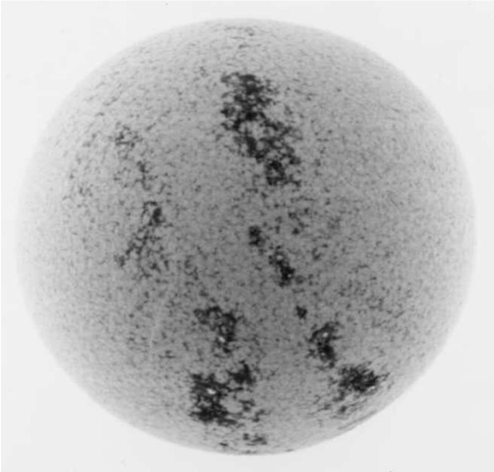

Fig. 3. Example of Calcium spectroheliogram taken on June 25th 1928 (negative image).

3. Calibration of the image

A procedure of calibration has been developped. We give here a summary and the details can be found on the web site.

1. *Dust and pit removal*: these images present notably some dust and pit which is important to remove (or at least reduce). To do this, a laplacian filter is applied to the image to locate discordant groups whose scale is smaller than the spatial resolution of the observing system. Typically 4,000 to 100,000 pixels out of 6.7 Mpix have their values replaced in this process.
2. *Initial guess of image center*: The solar image is identified in the 2601 by 2601 pixel array by first calculating the rms deviation of square sub-arrays of 5 by 5 pixels. Using the average of all these rms values as a dividing point, the centroid of all pixels having larger than average rms values is taken as the center of the solar image.
3. *Image geometry*: A strategy was adopted whereby the search in the image is for radii and image centers associated with the x (scan/dipersion direction) and y (cross-dispersion/slit direction) independently. Four separate locations are sought: x+, x-, y+ and y-. The center point in each direction is then the average of the + and - limb positions along each of these axes. With a tentative center, four quadrants of the image correspond to each of these limb positions. The expected time dependence value of the radius is used to constrain the solutions for the radii within bounds that are tightened successively during an iterative process.
4. *Intensity gradient determination*: The next step in finding the solar image limb is to calculate a gradient image using the first guess image center. Averages are calcultated for two 5 by 7 blocks of pixels which are symmetrically centered on the target pixel. The inner block average is substracted from the outer block average to form a smoothed measure of the intensity slope at each point.
5. *Reference gradient*: Due to exposure variations and optical system vignetting, the intensity and its gradient are highly variable from one image to another. In order to use the gradient images in spite of these variations, the first step is to determine typical values of the gradient for each quadrant by averaging the absolute value of the gradient in a strip of width 20 percent of the array dimension centered on the axes which go through the trial image center.
6. *Circular arc averaging - edge definition*: Still using the trial image center, a series of circular arcs can be defined as a function of trial image radii in each quadrant. The series of trial radii begins at a distance larger than the expected solar radius and is successively decreased until the gradient averaged along this arc segment exceeds 20 percent of the reference gradient.
7. *Scattered light correction*: Many of the images have scattered light primarily due to condition of the dispersion grating which was in a pit below the observing floor and subject to conditions of variable and high humidity. During each sequence of decreasing trial radius, the average of the

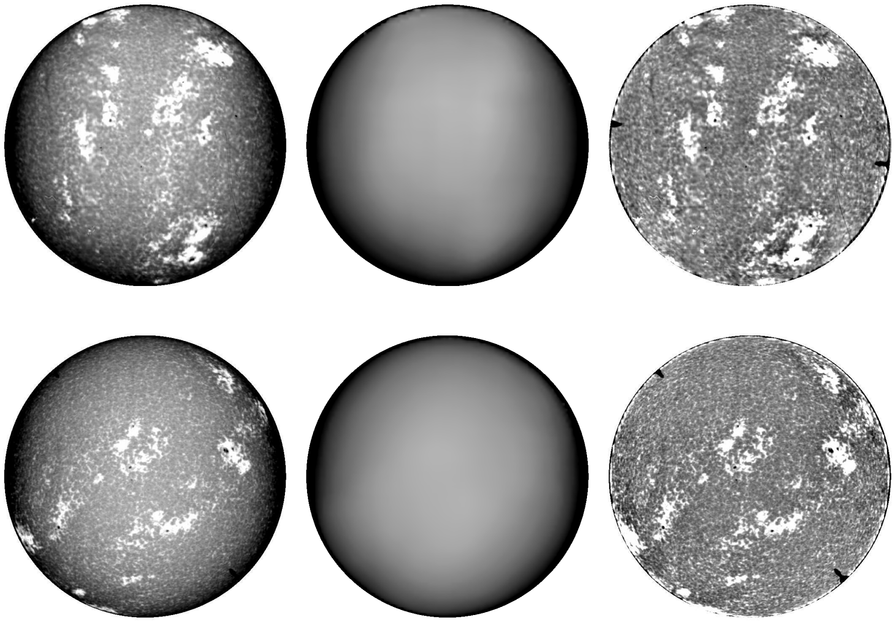

Fig. 4. Example of center-to-limb compensation for two images. Top: image taken on August 1st 1937; bottom: image taken on May 2nd 1939. From left to right: original image, extracted background image and flattened image. The roll angles have not been set to anything in particular and the pole position markers is indicated on each image.

gradients for the trial values more than two steps previously and less than 10 steps previously is calculated. This average is taken to be representative of the scattered light and is substracted from the gradient at the trial radius under consideration. The scattered light correction is limited to 50 percent of the reference.

8. *Veto conditions*: In addition to the basic criterion above, a number of veto conditions are tested and the trial radii will be decreased further if any of the veto conditions are found to be true. The veto conditions are: plate scratches produce paired and opposing gradients and plate edges produce strong gradients, which are either horizontal or vertical and which are not curved. If the gradient maximum is found to have one of these properties, it is identified as either a line (paired positive and negative) or an edge (unpaired) and is vetoed. Very strong dust spots or plate pits produce very high, paired slopes of opposite sign. If a dust spot is found along the averaging arc, the gradient maximum is vetoed. An area off the plate has a constant value equal to the clear scan result so that its rms is zero. Such areas are identified as blank and if the radius scan stops at a boundary of a blank area, it is vetoed. Finally, if the trial radius becomes smaller than 95 percent of the expected radius, the search is stopped. After the search is stopped, the test causing the previous point to be unacceptable is examined and its nature is used to calculate an image demerit value. The proper test to have failed is that the prior gradient was too small.
9. *Image center iteration*: After the radii are determined in the four quadrants, a new trial center is taken having its x and y coordinates as the averages of (x+ and x-) and

(y+ and y-) respectively. The quadrant radius searches are repeated iteratively until the trial center no longer changes.

10. *Image demerits*: After the sequence converges, a demerit value is computed based on veto conditions and the deviation of the image from its expected size and its deviation from being circular.
11. *Center-to-limb variation compensation*: One of the main problem we encoutered during the calibration process is the presence of a vignetting function. This function is linked to the centering of the optical system and more precisely to the relative position between the pupil and the grating. Due to that, the intensity and its gradient are higly variable from one image to another. Of course, it is important to compensate for the center-to-limb variation mainly due to the presence of this vignetting function. Several methods can be used. Here we use an iterative procedure with an adaptative median filter: we extract the background and divide the image by the background to obtain a flat image where the distribution of spots and faculae can be tracked to have access, for example, to the differential rotation rate. Figure 4 shows two examples of extraction of the background.

Other steps in the calibration, as the roll angle determination, are under investigation.

4. Future studies at UCLA

4.1. Solar differential rotation rate

The chromospheric network is clearly visible on a substantial fraction of the archive images from the CaK spectroheliogram sequence. Snodgrass and Ulrich (1990) have derived a rotation curve from the synoptic magnetogram sequence of the 150-foot tower. The CaK spectroheliogram sequence can be used in a similar fashion to provide a history of this process extending back to the 1915 era. If the pattern of differential rotation rate is in fact connected with the overall strength of the sunspot cycle, we should be able to detect this variability through the cross-correlation methods of Snodgrass and Ulrich (1990) applied to the CaK spetroheliograms sequence. Since the strength of the solar cycle has increased over the 20th century, any associated change in the rotation rate should be detectable.

4.2. Weak field magnetic fields

The strength of the widely distributed weak solar magnetic fields will be studied throught examination of the brightness of the CaK line emission over regions of the solar surface not directly associated with sunspots.

4.3. Total solar irradiance reconstruction

In collaboration with J. Pap from the Goddard Earth Sciences and Technology Center of the University of Maryland, the time dependence of sunspot and plage areas using the broadband images will be determined and the influence of these variations on the Total Solar Irradiance (TSI) will be estimated in order to reconstruct an improved history of the solar output energy.

5. Conclusions

As of June 2005, more than 20,000 CaK images spanning the period 1915 to 1952 have been scanned, identified with log book entries, analyzed for radius and image center and placed on the project web page as fits files. Vignetting and plate sensitivity functions are under investigation. Image roll angle needs to be determined and scattered light needs to be corrected. An analogous program with Arcetri solar archive is under progress at Rome Observatory under the responsablility of I. Ermolli and her team (Marchei et al. 2005).

Acknowledgements. This project is sponsored by NSF-ATM, AST and NASA-LWS and supported by the solar physics group at Stanford university.

References

Snodgrass, H. B. & Ulrich, R. K. 1990, ApJ, 351, 309
Marchei, E., et al. 2005, this volume