COMPARISON OF IMAGE-PROCESSING METHODS TO EXTRACT SOLAR FEATURES

L. Győri[1], T. Baranyi[2], M. Turmon[3], and J.M. Pap[4]

[1]Heliophysical Observatory, Gyula Observing Station, 5701 Gyula, P.O. Box 93, Hungary
[2]Heliophysical Observatory, 4010 Debrecen, P.O. Box 30, Hungary
[3]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, 91109, USA
[4]Goddard Earth Sciences and Technology Center, University of Maryland Baltimore County, Baltimore, USA

ABSTRACT

Large numbers of high precision solar images are now available from both terrestrial and space observatories, which has made it necessary to develop automated image processing techniques. In this paper we compare analysis of two sets of full-disk solar images: ground-based white light photographic films from Gyula and allied observatories, and magnetograms and quasi-continuum images from the Michelson Doppler Imager (MDI) on SOHO. We use two different automated image analysis techniques. The Sunspot Automatic Measurement (SAM) program has been developed at the Heliophysical Observatory at Debrecen for compiling the Debrecen Photoheliographic Data which are used to measure and catalogue the area and position of sunspots (umbra and penumbra). This project is part of the continuation of the Greenwich Catalogue. Startool is a general image analysis tool developed at the Jet Propulsion Laboratory, and applied to the MDI imagery as part of the SOHO Guest Investigator Program. As used with MDI, Startool identifies sunspots, faculae/network, and quiet sun using statistical pattern recognition techniques. Here we compare the area of sunspots as derived by Startool from the MDI images and by the SAM program as derived from the Debrecen and MDI images for the pilot interval of the second half of the year 1996.

Key words: sunspots, data processing.

1. INTRODUCTION

Observable solar features, such as faculae, network, and sunspots, are one manifestation of energy transfer processes in the Sun, and quantifying their characteristics can provide information about how these processes affect total and spectral solar irradiance, photochemical processes in the atmosphere, and other important phenomena. Areas of these features are used directly or as proxies in many fields of investigation, and they are obtained in several observatories with different observational and analysis methods. Precise area measurements are especially important for the studies of irradiance; for example, Fröhlich et al. (1994) found that one of the largest obstacles in irradiance modeling is the incorrect measurements of sunspot areas. Of course, the measured values of sunspot areas depend on the methods used to find them and the images themselves (Baranyi et al. 2001 and references therein). Therefore, it is important to compare the data obtained from various observatories and methods used to determine the areas of solar features, like sunspots.

In this paper we compare two full-disk image series. One of them is the SOHO/MDI images (1024^2 pixel) obtained as proxies for the continuum intensity near the Ni-I absorption line at 676.8 nm by combining the standard five filtergrams (Scherrer et al., 1995). These images are taken nominally 4 times a day, and processed by the MDI team to level 1.5. The MDI images we study have been further processed to remove various temporal and spatial artifacts as described by Turmon et al. (2002). The other image set contains daily ground-based photoheliograms taken to film or glass plates gathered from several observatories (Gyula, Debrecen, Kanzelhoehe, Kislovodsk), which are used in the compilation of the Debrecen Photoheliographic Data (DPD) catalogue (ftp://fenyi.sci.klte.hu/pub/DPD). We compare images during the interval 1996 July 1 – December 31, when there was overlap between the DPD and MDI data.

We apply two feature recognition techniques to these series. Startool is a general-purpose program that was developed to extract solar features from any group of temporally-synchronized images; we describe results for parameter settings derived for MDI quasi-continuum images and magnetograms. SAM has been developed to extract sunspot areas from white light ground-based observations for DPD.

2. IMAGE ANALYSIS METHODS

SAM is a set of cooperative computer programs that embraces every aspect of compiling a sunspot catalogue: from (1) setting up the necessary data basis for the observation and the telescope, (2) the measurements of sunspot
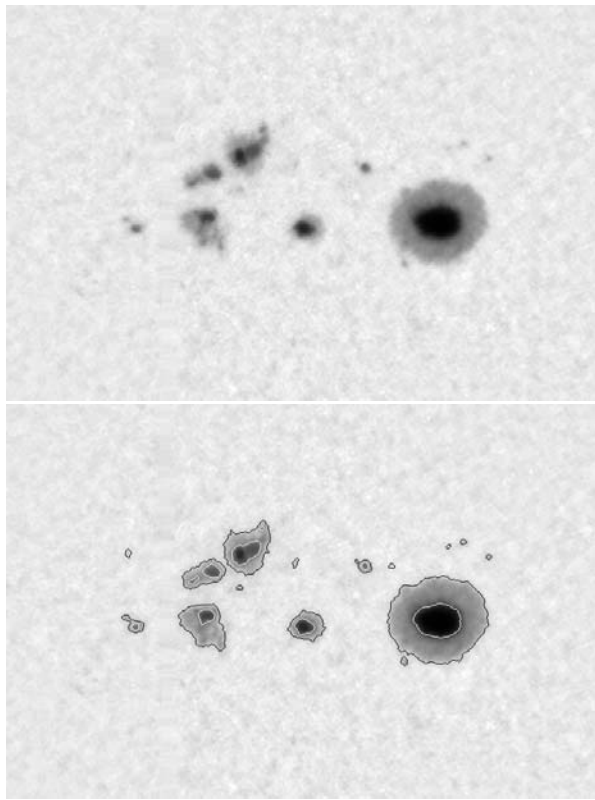
Figure 1: Active region NOAA 7981 on 1996 Aug. 2. Top: white-light image observed in Debrecen at 14:31 UT. Bottom: here we superimpose the boundary contours (penumbral is black, umbral is white) determined by SAM for the sunspots in the image.

coordinates on the solar disc and taking CCD image of sunspot groups through determining the heliographic coordinates of sunspots and measuring their area and (3) structural analysis of sunspots to making the catalogue ready for printing (Győri, 1998).

From the image, an iso-intensity contour set is produced so that every contour is determined at properly chosen intensity levels. The purpose of decomposing the image into a contour set is to get information about the relation between different solar features and their borders which are in this contour set (at least in a first approximation).

Using the gradient image, a gradient value is assigned to every contour as follows: the gradient values along the points of the contour are summed up and divided the number of the contour point. In what follows this gradient value is denoted by AGAC which stands for Averaged Gradient Along Contour. The contour set is ordered by the local minima (among these are the sunspots) of the image using the containment relation between contours, i.e., an ordered subset of the contour set is assigned to every local minimum. From the local minima, those contours which do not represent sunspots are filtered out. The penumbra border of a sunspot is the first (counted from the photosphere) contour having a local maximum in AGAC and the umbra border is the contour having the global maximum in AGAC.
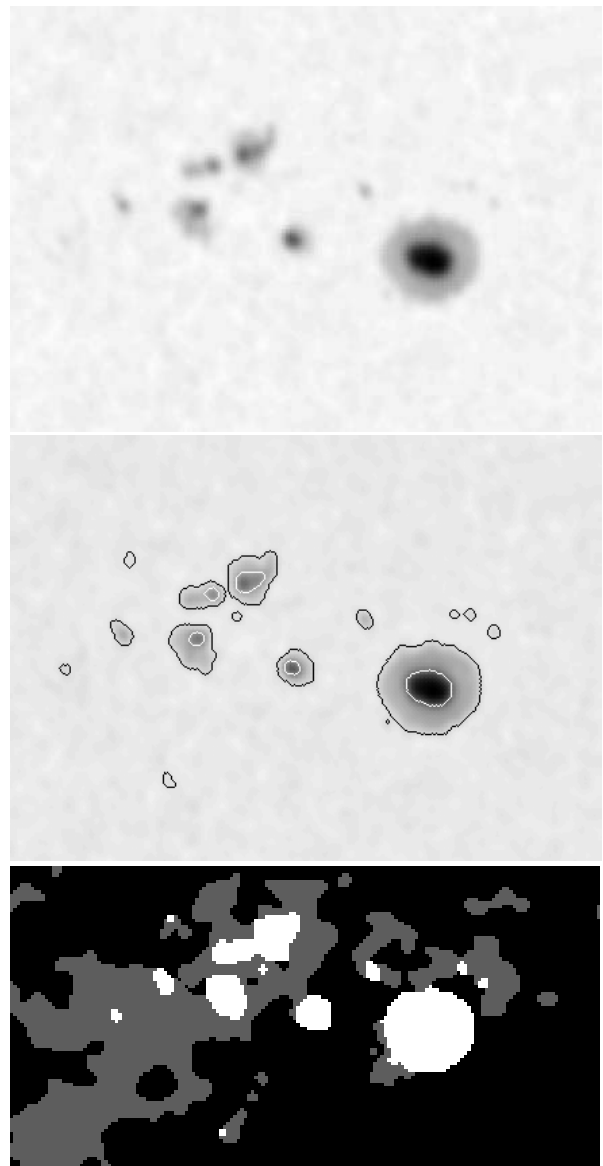

Figure 2: Active region NOAA 7981 on 1996 Aug. 2. From top to bottom: MDI quasi-continuum image observed at 14:48 UT, sunspot boundaries (penumbral, black line; umbral, white) determined by SAM and its labeling by Startool (white, gray and black indicate sunspot, faculae and quiet sun, respectively).

In certain cases, when different parts of a sunspot have different intensities, the border of the sunspot can not be represented by an iso-density line. To make allowances for that, the preliminary boundary obtained above is divided into intervals and the average gradient along an interval is compared to the values of the average gradient along similar intervals of the neighboring iso-intensity contours of the preliminary boundary and the interval of the preliminary boundary is replaced by the one having the maximum averaged gradient value.

An example of the sunspot (umbra and penumbra) boundaries is given in Figures 1 and 2. The outputs of SAM are the images of individual NOAA sunspot groups with the umbral and penumbral contours, and the projected areas

measured in pixels or in millionths of the solar disc (μd).

Startool is an image processing and analysis software package which allows automated, uniform, and objective analysis of long series of multi-observable images in a general-purpose framework (Turmon and Pap, 1997). The core function of this software is to label each pixel of the input images according to which type of activity is present there; other functions exist to preprocess images and interpolate them in time. The labeling method uses a statistical model, tunable for the specific application, to link pixel observables with feature types, and to indicate the pattern of local spatial dependence among pixel labels. Given the application-specific model, the software computes each labeling by maximizing its posterior probability given the observed images at that time.

In the MDI context, to separate sunspots and faculae, a two-parameter classification system has been applied, using both the magnetic field strength, as derived from the MDI magnetograms and the intensity values derived from the quasi-white-light filtergrams. This was done after finding that relatively weak magnetic fields (about 200 Gauss in MDI units) may cause either sunspots or faculae (Turmon et al. 1998). The model parameters described above are set by partly through scientist-provided labelings and partly through unsupervised clustering; the results described here are for the version 1 MDI model of June 2000 (Turmon et al. 2002). The outputs are full disk images (or *labelings*) which contain the extracted features (quiet sun, faculae, or sunspot) indicated by an integer tag — one means quiet sun, two indicates faculae, three for sunspot, as in Figure 2.

3. COMPARISON OF FEATURE AREAS

SAM is directed at finding sunspots in photograms, while the parameter settings for Startool discussed here find both sunspots and faculae using photograms and magnetograms together. To compare feature areas, therefore, we discard the Startool facular component and retain only the sunspots it identifies.

The two images series can be analyzed with either method yielding four types of sunspot areas:

- Applying SAM to the DPD photoheliograms i.e. DPD catalogue: $A_{\mathrm{DPD,SAM}}$.

- Applying SAM to MDI photograms: $A_{\mathrm{MDI,SAM}}$.

- Applying Startool to synchronized MDI photograms and magnetograms: $A_{\mathrm{MDI,ST}}$.

- Applying Startool to DPD photoheliograms.

Here we examine the first three cases.

3.1. Areas: MDI/Startool and DPD/SAM

The first step in the study is to compare sunspot areas from the two data sources to check their basic agree-
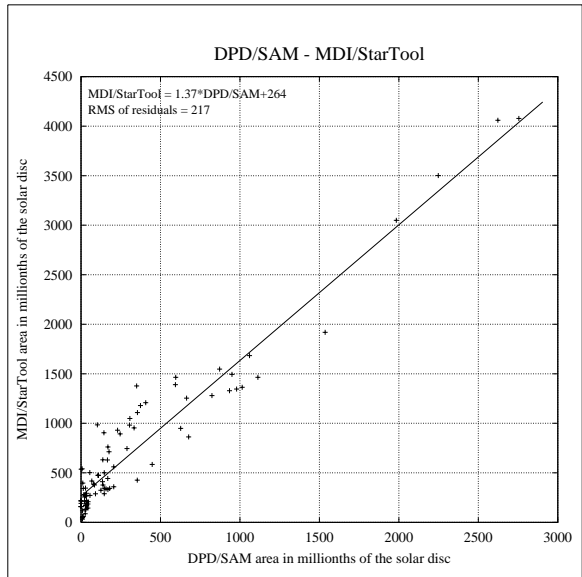

Figure 3: StarTool sunspot area from MDI images summed up on the whole disc versus DPD daily sums of projected areas from DPD photoheliograms .

ment. Since the DPD contains only one sum per day, we have chosen the MDI labeling closest to the corresponding DPD observation. Figure 3 shows the total area of sunspots on the whole disc obtained by SAM from the DPD photoheliograms and by Startool from the MDI images.

The equation and residual RMS of the regression line are:

$$A_{\mathrm{MDI,ST}} = 1.37 A_{\mathrm{DPD,SAM}} + 264 \qquad (\chi = 271). \quad (1)$$

As we can see, there are significant deviations between the sunspot areas derived from different observations with different methods. Especially remarkable is the disagreement at small areas (when DPD area is below 100 μd). The deviations can be attributed to several things, such as:

a. Differences in the two processing methods.

b. Differences in the solar images used by Startool and SAM.

c. Temporal differences between the two observations, resulting in evolutional and projectional differences.

The differences in the second item are mainly caused by three basic features. First, the white-light photographic images are taken inside the Earth's atmosphere. The MDI magnetograms and quasi-continuum images are taken in space, and are not subject to poor or time-varying seeing. Second, the MDI quasi-continuum images are created by combining five narrowband filtergrams, and are not true continuum images. Also, $A_{\mathrm{MDI,ST}}$ uses the MDI magnetograms which exist in two versions (1-minute and 5-minute integration times) having different noise characteristics. Third, the MDI images have a plate scale of 2 arcsec per pixel while the ground-based photographic
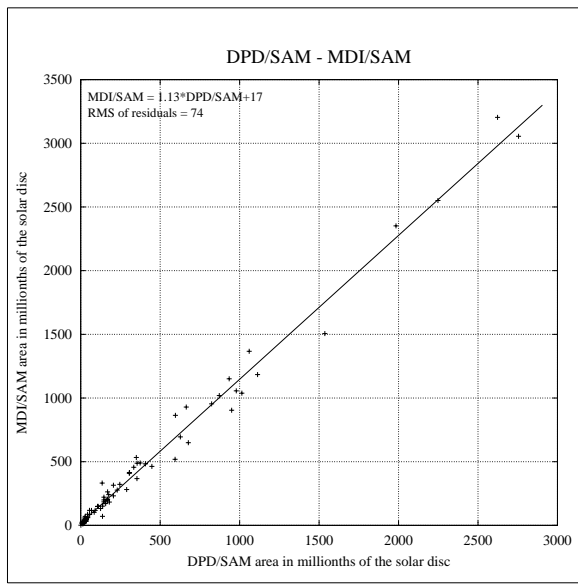
Figure 4: SAM sunspot area from MDI images versus DPD sunspot area from DPD photoheliograms.

Figure 5: Startool sunspot area versus SAM sunspot area from MDI images.

images have a spatial resolution of about 1 arcsec or better in case of good seeing. In case of the ground-based images we used a plate scale of 0.3 arcsec per pixel when taking CCD scans of these photographic images.

3.2. Areas: MDI/SAM and DPD/SAM

To separate and estimate the roles of the above effects, we have applied SAM to the MDI images. Figure 4 shows the total area of sunspots on the whole disc obtained by SAM from DPD photoheliograms and from MDI quasi-continuum images (discarding the magnetograms). The equation and residual RMS of the regression line are:

$$A_{\mathrm{MDI,SAM}} = 1.13 A_{\mathrm{DPD,SAM}} + 17 \qquad (\chi = 74). \quad (2)$$

We can see that there is a 13% systematic difference between the results when we apply SAM to MDI and DPD images, respectively, getting higher sunspot areas in the case of MDI. Cause (b) may explain this systematic difference, and (c) together with seeing effects on DPD may cause the scatter around the regression line. The residual deviations are strikingly smaller than in the case of $A_{\mathrm{MDI,ST}}$ versus $A_{\mathrm{DPD,SAM}}$ as we can see from the comparison of Figure 3 with Figure 4 and the corresponding equations. We conclude that, in Figure 3, the differences between the two image processing methods are the major factors causing the deviation.

3.3. Areas: MDI/SAM and MDI/Startool

Figure 5 shows the total area of sunspots on the whole disc obtained by SAM and Startool from MDI images. The equation of the fitted line is:

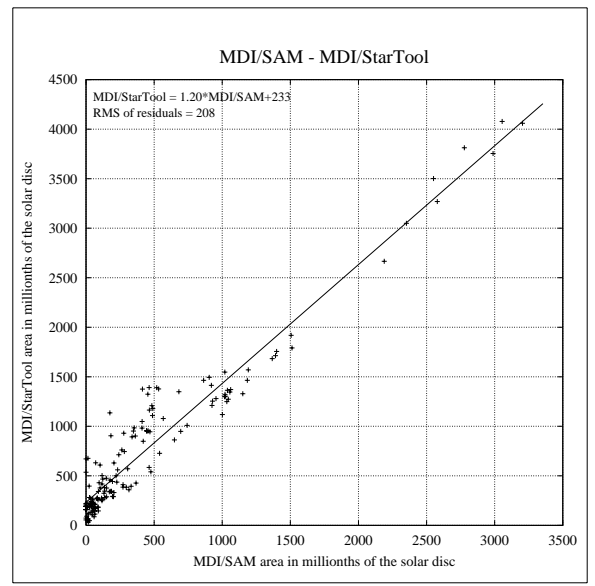$$A_{\mathrm{MDI,ST}} = 1.20 A_{\mathrm{MDI,SAM}} + 233 \qquad (\chi = 208). \quad (3)$$

As can be seen, the systematic deviation in this case is nearly half of that shown on the plot (Figure 3) in Section 3.1, where different methods were applied to different types of images. This decrease also indicates that some part of the deviations described in Section 3.1 can be attributed to the different image types. Here the same MDI images were processed with both methods, however, the deviations are larger than the ones described in Section 3.2, where the same method was applied to the two different types of images. This indicates that the larger deviations may come from the differences in the two processing methods.

3.4. Group areas: MDI/SAM and MDI/Startool

The relatively large Startool areas in Figure 3, when the DPD areas are below 50 μd, lead to the hypothesis that StarTool finds sunspot pixels outside SAM-identified sunspot groups. A direct comparison of the MDI images with the corresponding Startool-labelled images indicates that this may well be the case.

Figure 6 provides another argument in favor of this hypothesis. This figure was produced as follows: sunspot groups that were present in the DPD catalogue were cropped on the MDI and labelled images, and from these cropped images we determined the sunspot group areas. This operation excludes spot-groups which did not belong to a DPD sunspot group. We studied area data for 225 sunspot groups as derived by SAM and StarTool, respectively. The equation and residual RMS of the regression line are:

$$A_{\mathrm{MDI,ST}} = 1.17 A_{\mathrm{MDI,SAM}} + 82 \qquad (\chi = 149) \quad (4)$$

We can see from the figure and from the parameters of the regression line that in this case we get a better agreement between the areas derived by SAM and StarTool from the
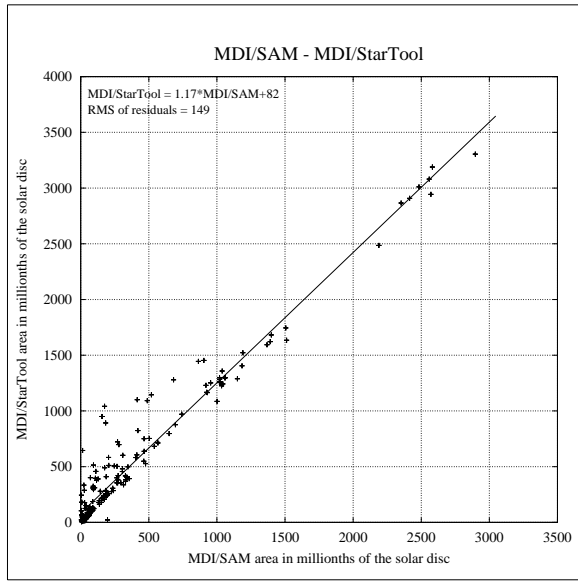
Figure 6: Startool versus SAM sunspot area from cropped labelled and MDI images.
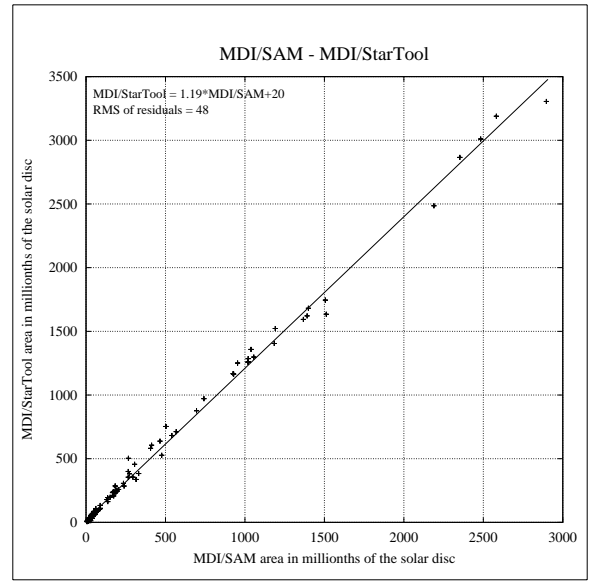

Figure 8: Startool versus SAM sunspot group areas from cropped MDI and labelled images. The distances of the sunspot groups from the center of the solar disc are smaller than 0.7 relative radius.

MDI images. This shows that the largest part of the deviations are associated with pixels which have not been identified as sunspots by SAM.

Figure 7 shows a Startool labelled image (black and gray indicate sunspot and quiet sun, respectively). To see the sunspots more easily, we omitted the identified faculae from this image. The MDI image, to which Startool was applied to, was taken on December 3, 1996 at 10:59:35 UT. According to SAM (either applied to the MDI image or the Debrecen photoheliogram) there was no sunspot on the Sun this day. We can raise the question whether these small features found by Startool are actually sunspots. To answer this question, we need to study higher resolution images than provieded by MDI.

If we look at the MDI and the labelled images near the solar limb we can see that some facular pixels are identified as sunspot by StarTool. This seems to stem from the under-representation of such high-contrast faculae in the data selected to fit the model which drives Startool. To reduce this known cause of systematic error, we omitted the sunspot groups near the limb. Figure 8 depicts SAM and Startool sunspot group areas from cropped MDI and labelled images. In this subset, sunspots are located within 0.7 relative radius from the disk center. The regression line is:

$$A_{\mathrm{MDI,ST}} = 1.19 A_{\mathrm{MDI,SAM}} + 20 \qquad (\chi = 48). \quad (5)$$

As can be seen, the RMS of residuals decreased and the data fit to the regression line very well. However, a systematic difference of about 19% still remains, which most probably arises from the differences in the two image processing methods.

4. CONCLUSIONS

The results presented in this paper point out the differences in sunspot areas as measured from different images and by different image processing techniques. Specifically, we have found that the application of the "Sunspot Automatic Measurements" (SAM) software to the Debrecen Photoheliographic Data (DPD) and the MDI quasi-continuum images gives about 13% larger MDI spot areas than DPD areas. Part of this discrepancy between the two spot area measurements may be related to the fact that the images were taken at different times. However, most of the discrepancy due to observation time differences (two-dimensional projection effects as well as sunspot evolution) should show up the RMS of the residuals of the linear fit of the areas. Therefore, one can assume that much of the systematic difference is due to differences between the DPD and MDI images.

We have also found that when we apply SAM and Startool to the MDI quasi-continuum images, the two image processing programs classify features differently. For example, Startool identifies pixels distributed over the solar disk as sunspots which are not recorded as spots by SAM. Also, close to the limb, Startool sometimes identifies facular pixels as sunspot pixels due to a model incorporating too few high-contrast faculae. Even if we take into account these apparent differences in feature identification by the two image processing programs, the MDI sunspot areas measured by Startool are systematically larger by about 19% than measured by SAM. These results clearly demonstrate the necessity of intercomparison of various images and image processing techniques to derive the sunspot areas as correctly as possible. The results also indicate that we need to study higher resolution images than provided by MDI to better account for the sunspot effect on solar irradiance variations and therefore to im-
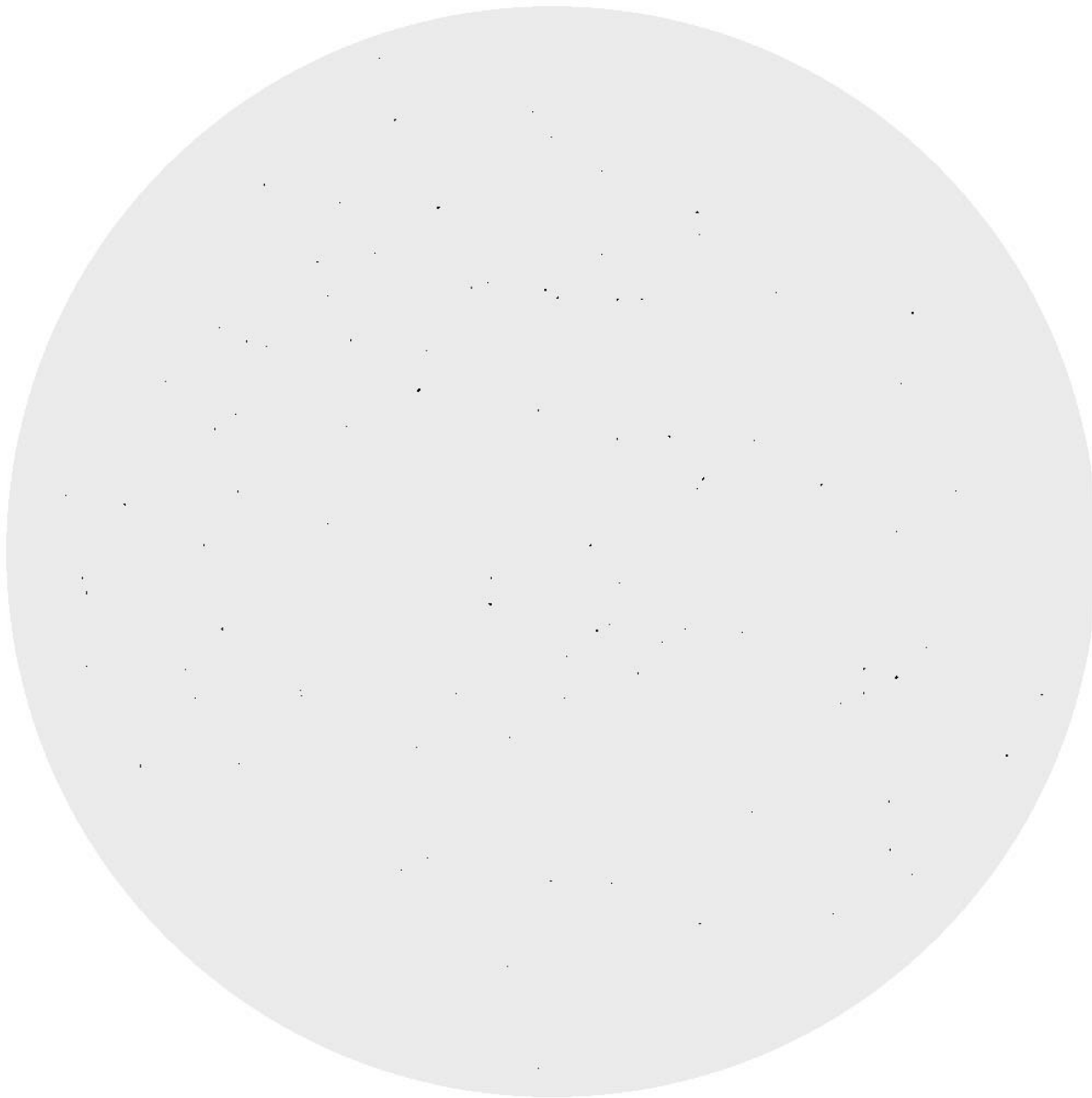
Figure 7: Startool labelled image (black for sunspot, gray for quiet sun and faculae) derived from MDI images taken on 1996 Dec. 3 at 10:59:35 UT.

prove irradiance models.

REFERENCES

Baranyi T., Győri L., Ludmány A., Coffey H.E., 2001, MNRAS 323, 223

Fröhlich C., Pap J.M., Hudson H.S., 1994, Solar Phys. 152, 111

Győri L., 1998, Solar Phys. 180, 109

Scherrer P.H. et al. 1995, Solar Phys. 162, 129

Turmon M., Pap J., 1997, Statistical Challenges in Modern Astronomy II (eds. G.J. Babu and E.D. Feigelson), Springer-Verlag, 408

Turmon M., Pap J.M., Mukhtar S., 1998, Structure and Dynamics of the Interior of the Sun and Sun-like Stars, ESA SP-418, 979

Turmon M., Pap J.M., Mukhtar S., 2002, ApJ, 568, 396