Deliverable 2.2:
SOHO/MDI Sunspot Data

Project acronym: *SOTERIA*
Project title: *SOlar-TERrestrial Investigations and Archives*
Grant Agreement number: *218816*
Coordinator: *Giovanni Lapenta*
Project co-funded by the European Commission,
Seventh Framework Programme
Funding Scheme: *FP7-SPACE-CALL-1*

Due date of deliverable:	**Month 36, October 31, 2011**
Actual submission date:	**10/12/11**
Start date of the project:	**Nov 1, 2008**
Project duration:	**3 years**

Work package:	**2**
Task(s):	**2.2SOHO/MDI Sunspot Data (SDD)** **2.3White-light measurements** **2.4Magnetic observations** **2.5Standardisation of database formats**
Lead beneficiary for this deliverable:	**KO**
Editor:	**A. Ludmány**
Authors:	**Győri, L., Baranyi, T., Ludmány, A.**
Quality reviewer:	

Project co-funded by the European Commission within the Seventh Framework Programme (2007)		
Dissemination level		
PU	Public	PU
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

DOCUMENT CONTROL

This document has no controlled or maintained paper copies. The master document is held in CPA servers on the KULeuven network. All files have widespread access for reading only. Changes to this document will be notified to its audience.

ISSUE RECORD

Version	Date	Author(s)	Reason for modification	Status
1.0	10/04/11	Győri, L., Baranyi, T., Ludmány, A.	First version	Public distribution

Report on the on-line publication of the SOHO/MDI sunspot data

Type of the deliverable:
Public on-line publication of a database.

The SOHO/MDI sunspot data (SDD) are accessible at:

http://fenyi.solarobs.unideb.hu/SDD/SDD.html

Content of the database:

1. Formatted tables of sunspot area, position, and mean magnetic field data derived from SOHO/MDI observations- ASCII

Type of datasets: dataset of sunspot s without gathering into sunpot groups (fdSDD*YEAR*.txt), combined dataset of sunspots groupped into NOAA sunspot groups (SDD*YEAR*.txt), separeted tables for daily sums of sunspot areas (dSDD*YEAR*.txt), total areas and mean positions of sunspot groups (gSDD*YEAR*.txt), and individual sunspot data (sSDD*YEAR*.txt).
The following data are included into numerical catalogue in text format: date and time of observation, positions of sunspots, mean position of sunspot groups (Carrington and polar coordinates), umbral and penumbral areas (both projected and corrected), mean magnetic fields of umbrae and penumbrae, total areas and mean positions of sunspot groups. The NOAA numeration of the groups is checked and in case of necessity completed or corrected.

2. Quality-filtered and time-filtered full-disk intensity images selected from Hourly Data Sets Level 1.8.– 16-bit FITS, 8-bit JPG
3. Quality-filtered and time-filtered full-disk magnetograms selected from the recalibrated Hourly Data Sets and Daily Data Sets Level 1.8.– 32-bit FITS, 8-bit JPG
4. Intensity images of sunspot groups – 16-bit FITS, 8-bit JPG
5. User-friendly graphical presentation of the data – HTML
6. User-friendly MySQL query for SDD – HTML
7. Formatted tables of sunspot group tilt angles derived from SDD – ASCII

Time resolution:
~ 1 hour depending on the actual time resolution of the SOHO/MDI observations

Spatial resolution:
Original observations:1k x 1k pixels
Processed images:3k x 3k pixels

Time interval:
1996-05-19 - 2010-12-31

The related Debrecen Photoheliographic Data (DPD) are accessible at:

http://fenyi.solarobs.unideb.hu/DPD/index.html

Motivation of the deliverable:

The sunspots, the most important and longest observed features of solar activity have been documented with much less detail than necessary and possible for a very long time. The earlier datasets are very important but insufficient. The International Sunspot Number is the longest and simplest dataset. The Greenwich Photoheliographic Results (GPR) is the longest sunspot catalogue but it disregards individual sunspots in the majority of its volumes. The Debrecen Photoheliographic Data (DPD) is the first catalogue containing position and area data of all observable sunspots and sunspot groups using ground-based observations. However, a catalogue of detailed sunspot and sunspot group data including magnetic data with time resolution higher than one day was missing before the SOTERIA project.

Progress beyond the state of the art:

The observational material of the SOHO/MDI instrument yielded the first opportunity to compile a really complete sunspot dataset which made advancements in the following respects.

1. The cadence of the data is about 1 hour (minimum of time difference is 30 minutes, maximum depends on the actual time resolution of the SOHO/MDI observations). SDD contains data of all sunspots and sunspot groups, even the smallest ones with this high time resolution. All earlier datasets were made on a daily basis.

2. This is the first sunspot catalogue which contains magnetic field data of sunspots besides the position and area data.

3. A new user-friendly html presentation is developed for SDD and DPD. Its features are:

- Numerical data of sunspot and sunspot groups,
- Quick look of the sun's view in both white light and magnetogram with links to the active regions, the user can turn the pages of the observations and follow the progress of activity,
- Separate images of active regions with the numeration of spots referring to the numerical data.

4. A new on-line MySQL tool was develped to search data of specific regions in the numerical datasets.

5. A new service has been launched: a quick-look catalogue has been published whith all kinds of information on a daily basis updated immediately after receiving the latest observations. This is the most detailed up-to-date data source about the solar photospheric activity.

6. The catalogue of active region tilt data derived from SDD is the first dataset which contains tilt data with two new different tilt definitions including the information of magnetic polarity of spots besides the traditional definitions. This part of D 2.2 exceeds the goals formulated in Annex I.

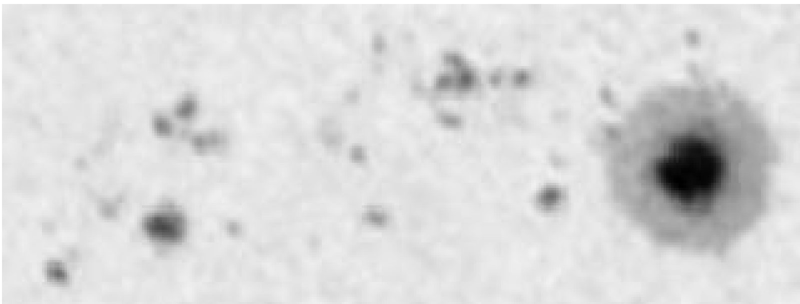
The work pursued:

The software package named SAM (Sunspot Automatic Measurement) (Győri, 1998) was upgraded, improved and implemented for the purposes of the SDD work. The pairs of continuum observations and magnetograms was selected closest in time within 50 minutes time interval. The different sets of parameters of SAM were tested, and the results were compared with the data derived from the high spatial resolution observations of DPD.The software run with that set of parameters which derived the data closest to the data of DPD. The sunspot borderlines were checked by visual inspection, because the small spatial resolution of MDI images caused some distortion of borderlines in some cases. This was the most manpower-requiring phase of the process. The damaged pixels also needed special treatment. After the separation of sunspot groups, the checked data were edited into the numerical tables and the relevant html-pages were produced. The production and revision of the DPD data was a parallel procedure, because the DPD data served as needed input data to derive the SDD data from the MDI images and to check the data in regular cross-comparison and data validation. There was a previous version of SDD for 1996-97 at the baseline, but the whole 1996-2010 interval was reprocessed with the new magnetograms recalibrated at the end of 2008.

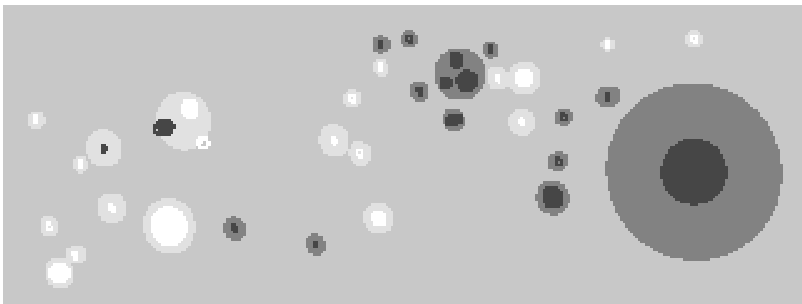
The quality of the new catalogue:

The completeness and precision of the data can be demonstrated graphically in the following figure showing the active region NOAA 10865, 1.Apr.2006, 01:35:32UT. The first and third frames show the SOHO/MDI continuum image and magnetogram, the middle frame shows the drawing of the active region reconstructed from the catalogue. The sizes of the umbrae and penumbrae of the spots are represented by appropriate circles, the colours represent the magnetic polarities. All important properties are documented in the catalogue.

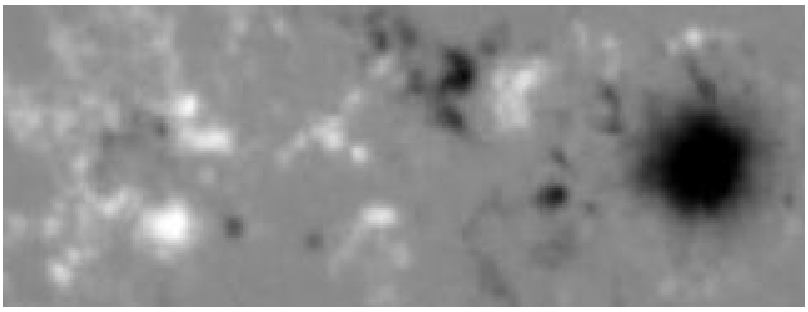
continuum image



reconstructed data



magnetogram



Comparison of SDD data with SDO/HMI data:

The SDO solar images have much higher resolution (~0.5"/pix) than the SOHO solar images (~2"/pix). This difference in image resolution has a significant impact on the sunspot areas measured in the images. The comparison area of sunspots derived from the SDO/HMI and the SOHO/MDI images helps in the calibration of the SOHO sunspot area to the SDO. For the comparison 1750 image pairs, each within 2 minutes, between May 1 and December 31 in 2010, were chosen. The relationship between sunspot data is modelled by linear regression where the dependent variable is the HMI area.

The comparison of the HMI and the MDI total spot area is summarized in Table 1 where slope and y-intercept are parameters of the regression line, RRMS is the Residual Root Mean Square, and corr. coeff. is the correlation coefficient obtained from the data. As we can see, the MDI total spot areas are larger than the HMI in all the three cases, however, the parameters of the regression lines are different. The trend is that at larger areas the slope is lower and the y-intercept is higher.

Table 1. Results of the total spot (umbra+penumbra) area comparison

	slope	y-intercept (md)	RRMS (md)	corr. coeff.
all data	1.06	5	21	1.00
area <= 300	1.07	2	13	0.99
area > 300	1.03	28	32	0.99

The comparison of the HMI and the MDI total umbra areas is summarized in Table 2.
As we can see, the MDI total umbra areas are smaller than the HMI in all the three cases, however, the parameters of the regression lines are different . The trend is that at larger areas the slope is higher and the y-intercept is smaller.

Figure 1 shows the borderlines of sunspots in MDI and HMI images, and it illustrates the causes of the differences.

Table 2. Results of the total umbra area comparison

	slope	y-intercept (md)	RRMS (md)	corr. coeff.
all data	0.94	-3	9	0.98
area <= 50	0.95	-2	6	0.90
area > 50	1.01	-12	11	0.97

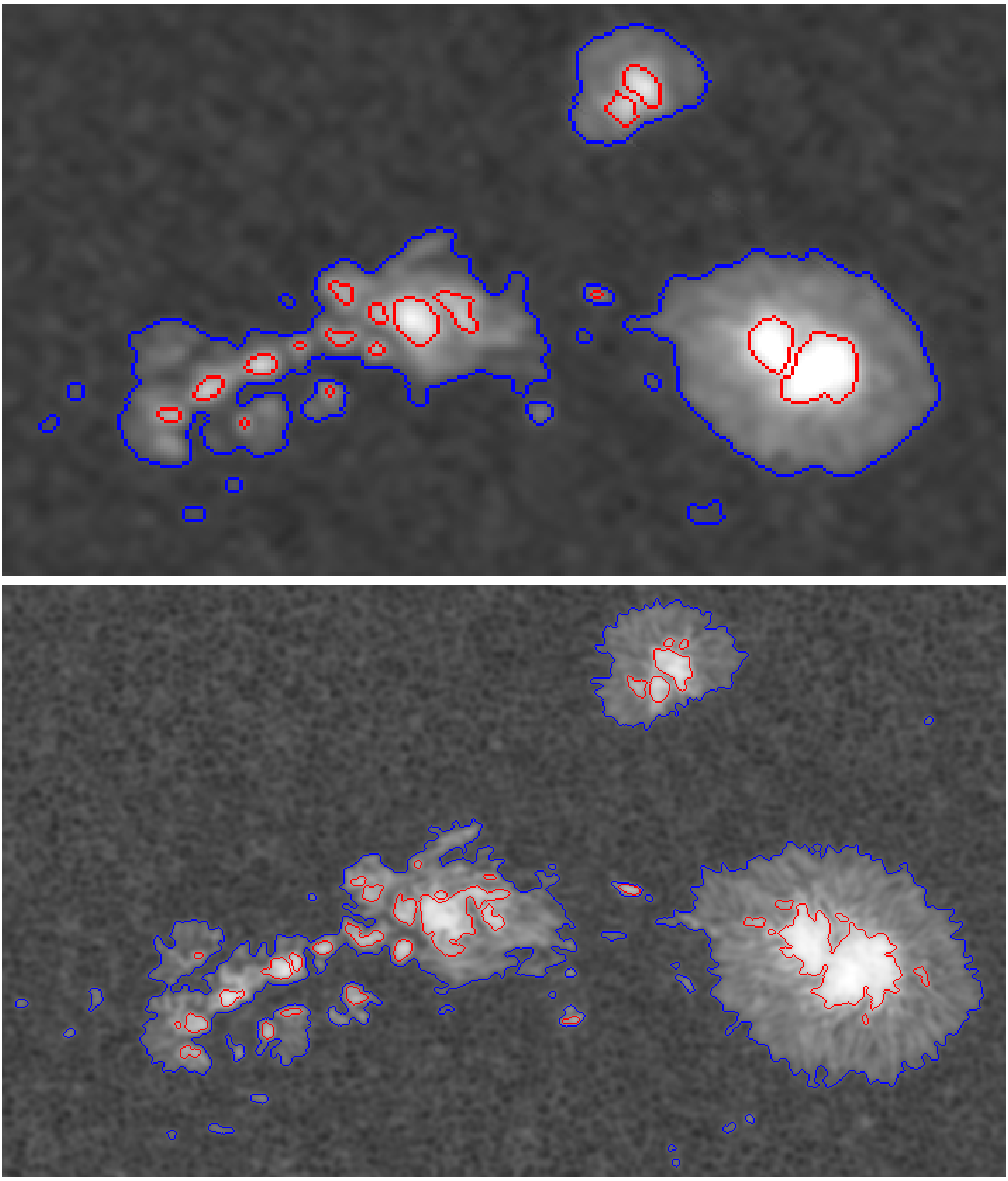


Figure 1. Upper panel: Sunspot group in a negative MDI intensity image. Lower panel: The same sunspot group in a negative HMI image in the same time. Blue line: borderline of penumbra, red line: borderline of umbra.

It can be seen in Figure 1. that due to the better resolution of the HMI images, even in smaller sunspots, the umbra separates away from the penumbra background, and also in large penumbrae, even the small umbrae separate away from the penumbra background. Thus they can be perceived as it can be observed in the HMI image, and this is why that total umbra area are generally larger in HMI images than in MDI ones.

In looking for the explanation of differences in solar feature areas in image sets with different resolutions, three causes are to be distinguished:

a) Border thickness. SAM finds the the border contour of the solar feature and from it determines the area of that feature. In the case of the HMI and MDI images, one MDI border pixel represents 16 times more area on the solar disk than one HMI pixel. If we suppose that the true border halves the pixel then the contribution of one MDI border pixel to the feature area is 8 times of one HMI border pixel. The overall effect of the border thickness on area depends very much on the raggedness of the border. This effect makes the feature area larger for lower resolution images.

b) Structural adequacy. How well the real structure of a solar feature is mirrored in a solar image depends on its resolution, for example, fissures separating nearby solar features or making their border irregular do not show up in a lower resolution image and so their areas are included into the area of the feature. Similarly, small nearby features merge in a lower resolution image. This effect also makes the feature area larger for lower resolution images.

c) Feature perceivability. Features that are small (below or around one pixels) and less intensive desolve into their enviroment and so they can not be perceived. This effect makes the total feature area larger for higher resolution images, and HMI images have about 3 times more spots (umbra + penumbra without umbra) and umbrae than MDI.

Conclusion:

The low spatial resolution of MDI images affects the area data of sunspots in such a way that the MDI total spot areas are larger and the MDI total umbra areas are smaller than the related areas derived from the high resolution HMI images. However both the slope of the regression line and correlation coefficient are close to 1 showing that the difference is statistically small. This means that the SDD contains high quality data suitable for new studies of high precision.

New perspectives.
The new catalogue opens the possibility to examine the development and morphology of the solar active regions in full detail, high temporal resolution and large statistical sample. This will give opportunity to study, among others, the dynamics of emergence/decay and the formation of unstable magnetic configurations, possible sources of explosive events.

Reference:
Győri, L., Automation of area measurement of sunspots, Solar Phys., 180, 109-130, 1998