

In-depth survey of sunspot and active region catalogs

Laure Lefèvre¹, Frédéric Clette¹ and Tunde Baranyi²

¹Royal Observatory of Belgium, 3 Avenue Circulaire, 1180, Uccle, Belgium
email: laure.lefevre@oma.be and frederic.clette@oma.be

²Heliophysical Observatory of the Hungarian Academy of Sciences, H-4010 Debrecen, Hungary
email: baranyi@tigris.unideb.hu

Abstract. When consulting detailed photospheric catalogs for solar activity studies spanning long time intervals, solar physicists face multiple limitations in the existing catalogs: finite or fragmented time coverage, limited time overlap between catalogs and even more importantly, a mismatch in contents and conventions. In view of a study of new sunspot-based activity indices, we have conducted a comprehensive survey of existing catalogs.

In a first approach, we illustrate how the information from parallel catalogs can be merged to form a much more comprehensive record of sunspot groups. For this, we use the unique Debrecen Photoheliographic Data (DPD), which is already a composite of several ground observatories and SOHO data, and the USAF/Mount Wilson catalog from the Solar Optical Observing Network (SOON). We also describe our semi-interactive cross-identification method, which was needed to match the non-overlapping solar active region nomenclature, the most critical and subtle step when working with multiple catalogs. This effort, focused here first on the last two solar cycles, should lead to a better central database collecting all available sunspot group parameters to address future solar cycle studies beyond the traditional sunspot index time series R_i .

Keywords. Catalogs, surveys, sun: photosphere, sunspots, methods: data analysis, statistical

1. Introduction

So far, the main sunspot time series available for research has been the International Sunspot Index R_i obtained from a large number of observatories and visual observers. Until the last decade, the limitations in digitization equipment, in image processing techniques and computing capacity largely prevented building more detailed sunspot series. However, the current advances in theoretical modeling require additional long-term observational constraints and we now finally have appropriate means to create more complete long-duration catalogs.

Previous research based on catalog sets focused on single or just a few parameters for more detailed studies, often dropping the rest of the information (Balmaceda *et al.*, 2009). As an example, in an earlier comparative study of the discrepancies between different measurement of sunspot areas (Gyori & Baranyi 2006, Gyori *et al.*, 2005), the scope was limited to combining areas and sometimes also the corresponding latitudes and longitudes, thus leaving out other sunspot data.

The purpose of the work presented here is to extend the exploration to the complete parameters-space over extended periods of time. We considered a base of 20 catalogs. Most of them can be found on the NGDC website (see bibliography). To initiate this

study, we use two particularly rich catalogs. The best choice for a multi-parameter study of sunspot activity is the Debrecen Photoheliographic Data catalog (Mezo & Baranyi, 2005, Gyori *et al.*, 2005). It is well complemented by the USAF multi-station catalog, which combines data from a worldwide network of solar observatories, mostly run by the US Air Force. We will now describe in more detail our comparative analysis of those catalogs.

2. The Debrecen and USAF data

The DPD catalog covers the last two solar cycles and is still in construction and the digitization work is still in progress in the framework of the *SoTerIA* project (Lapenta, 2007). Most of the DPD catalog is based on photographic and CCD images from the *Gyulia and Debrecen Observatories*. However, for the period from 2004 to 2006, in order to fill in promptly the catalog in support of this study, SoHO images were used instead: they do not offer the same level of details, but their preprocessing is faster (Baranyi T., private communication). *USAF* observations come from 9 stations between Greece and Australia. The sunspot group entries for each day are based on a maximum of 6 of those observatories.

Both datasets differ by 3 main aspects : (1) the time coverages are different: 86.3% for the USAF series versus 98.9% for the DPD series. (2) The USAF catalog only lists groups while the DPD catalog lists individual sunspots, with grouping added only afterwards. (3) The USAF catalog is structured according to the evolution of sunspot groups, while the DPD adopts a chronological ordering.

In order to assess the consistency of these datasets, we computed the standard *Wolf Number* (W_{OBS}) with the sunspot and sunspot group counts extracted from the catalogs. We compared W_{USAF} and W_{DPD} to the International Sunspot Index (R_i), and Wolf Numbers from several individual stations : Catania (CA), Locarno (LO), Kanzelhöhe (KZ) and Uccle (UC). *Figure 1* plots the resulting series.

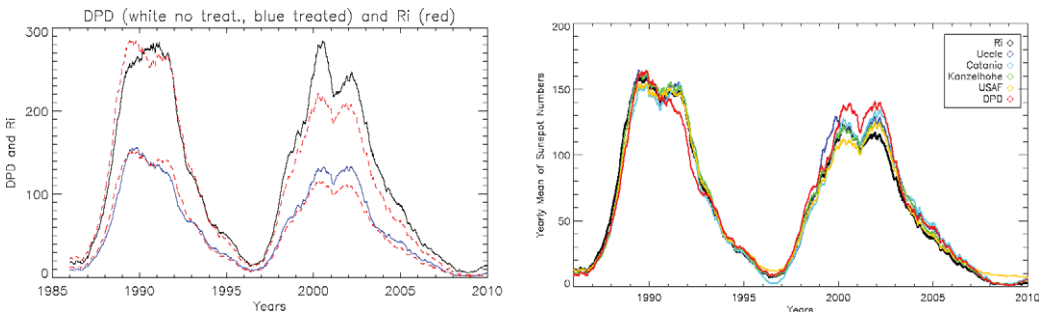


Figure 1. On the left, the Wolf number derived from the DPD catalog, W_{DPD} , from 1986 to 2010 (one-year running mean of daily values) is plotted for the full data set (black) and for the reduced set after filtering as described in the main text (blue). Each series is compared respectively with the R_i index (lower red dashed line) and with $R_i/0.6$ (upper red dashed line). On the right, W_{DPD} after filtering (red curve) is compared with several independent sunspot number series: R_i (SIDC, black curve), Uccle, Catania, Kanzelhöhe and the USAF catalog

While W_{USAF} agrees closely with other series, W_{DPD} shows significant deviations. The raw DPD values shown in Fig. 1 (left panel) give almost the same peak value for

cycles 22 and 23 ($max_{22}/max_{23} = 1.08$), while all other series consistently show that #22 was significantly higher than #23, with ratios of 1.38, 1.36, 1.33, 1.29 and 1.17 for W_{LO} , R_i , W_{USAF} , W_{UC} and W_{CA} respectively. Although there is an obvious scatter in the values, the systematic character of this discrepancy points at a significant bias in W_{DPD} . Moreover, the cycle #22 maximum does not show the characteristic double peak present in all other series.

The cause of this difference resides in the details of the DPD catalog, which lists about 10% more sunspot groups than, e.g. the USAF catalog. After checking the characteristics of small sunspots listed in the DPD catalog, we see that approximately 50% of the spots are classified as penumbra without umbra and their status as spots should be carefully reconsidered. This seems to be true also for spots with a diameter smaller than 2 arcsec.

This R_i reconstruction indicates that the DPD catalog is highly comprehensive to the point of including some features that do not qualify as sunspots in long-term visual observations. After this filtering, W_{DPD^*} matches better R_i and other individual series (Fig. 1, right panel). The ratio between maxima rises to 1.17, which is still on the low end but equivalent to the Catania ratio (Figure 1). However maximum #23 is still higher than in all other series and the maximum of cycle #22 still does not reproduce well the second peak.

3. Merging catalogs

After exploiting this sunspot-level selection only possible with the DPD catalog, as the DPD and USAF catalog contain different information about sunspot groups, we can take advantage of the time overlap between both catalogs to merge the contents, e.g. adding the USAF Zürich modified morphological classification (Zpc) to the DPD sunspot group catalog. In order to obtain a closer match in observing times, we first reorganized the USAF catalog entries by selecting the station observing at the time closest to the DPD time. We then matched each individual group listed in both catalogs for each date, using the Euclidian distance between groups with differing NOAA group numbers. In this process, the main difficulty came from the difference in the group splitting methods implemented in those two catalogs. This was further complicated by internal ambiguities in the respective group splitting and numbering schemes when the group evolution over successive days led to the insertion of extra "orphan" groups. By a lack of standardization, such groups are intercalated by adding suffixes to the official NOAA group number. This leads to a confusion between those sub-designations, e.g. NN for NOAA, becoming NN , NN_a or NN_A in either the DPD or USAF catalog. The group identification program must then extend the search to all possible nomenclatures in both catalogs.

Overall, *about 80% of the DPD sunspot groups have a straight forward correspondence in the USAF group catalog. The remaining 20%, where the matching is problematic, are mainly related to groups from the DPD catalog that do not appear in the USAF catalog (11%), because of the higher level of details in the DPD catalog. They are also largely related to cases where the USAF and DPD groups are too distant to be matched (5%).*

As illustration, Figure 2 shows the sunspot regions for May 30th 1991 for DPD, USAF (black and red circles) and Kanzelhoehe data. Three groups (6655, 6649, 6653 and 6653c) are listed only in the DPD catalog, suggesting that the latter is more comprehensive and includes spots that escaped detection in the other observations. Moreover, the size of the sunspot groups in the DPD data appears systematically larger than in the USAF data (see group 6644, fig 2. RIGHT panel), thus reflecting again the difference in standards and methods between catalogs of different origins and built at different epochs. Finally,

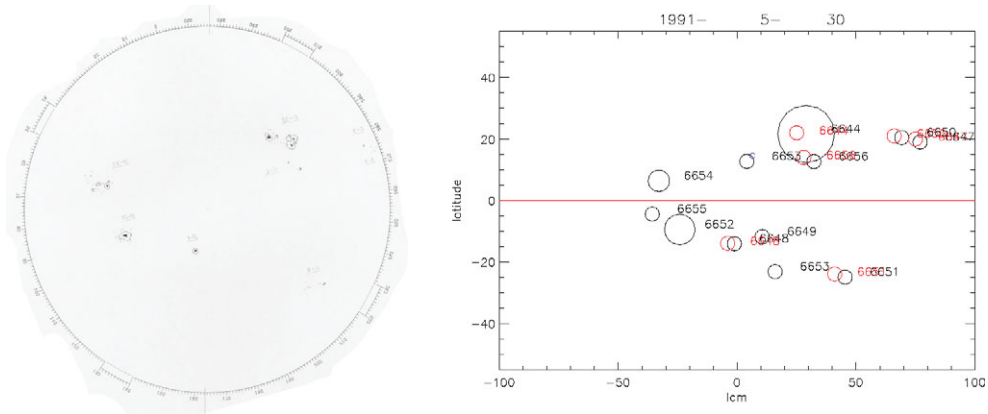


Figure 2. The left panel represents the drawing from Kanzelhöhe Observatory for May 30th 1991 while the right panel presents a synoptic map of the surface of the Sun with DPD (black) and USAF groups (red) for the same day. The size of the circles is proportional to the total area of the considered group.

groups 6652 and 6654 appear in the Kanzelhöhe and Debrecen data on May 30th but only on June 1st for the San Vito Observatory (USAF). Such mismatches can be caused either by the difference of sunspot grouping schemes or by a difference in observing times (date jump). Solving such cases often requires additional external information not necessarily included in other existing sunspot catalogs. This illustrates the work currently in progress and that will ultimately allow a full merging of those two catalogs, after removal of the unsolved ambiguities.

4. Conclusion

Presently, we have produced a first version of this extended composite catalog spanning the last two solar cycles, from 1986 until now. It contains groups, spots, areas and Zürich modified morphological types, as well as a basic computing of the extent of the groups. The catalog will also be enhanced by the addition of extra parameters derived from the base ones (e.g. dipole tilt angles) or extracted from other databases or catalogs, where they overlap the entries of this catalog. The resulting catalog will be made available through the SIDC website (www.sidc.be). Beyond the rich but time-limited DPD and USAF catalogs exploited here, we will then work backward in time by adding other older or sometimes longer catalogs. However, these catalogs contain less sunspot information and also pose more problems of standardisation, making the merging and group matching more complex. It will thus be necessary to implement advanced techniques like Multivariate Data Analysis (Murtagh F. & Heck, A., 1987). This effort will build on the knowledge acquired during this first test-bed offered by the DPD and USAF catalogs.

References

- NGDC website: <http://www.ngdc.noaa.gov/>
 Balmaceda, L. A., Solanki, S. K., Krivova, N. A., & Foster, S. 2009, *Journal of Geophysical Research (Space Physics)*, 114, 7104
 Gyori, L., Baranyi, T., Muraközy, J., & Ludmány, A. 2005, *MmSAI*, 76, 985
 Gyori, L., Baranyi, T., Muraközy, J., & Ludmány, A. 2005, *MmSAI*, 76, 981

- Györi, L. & Baranyi, T. 2006, *SOHO-17. 10 Years of SOHO and Beyond*, 617,
Lapenta, G., SOTERIA Team 2007, *AGU Fall Meeting Abstracts*, A338
Mezo, G. & Baranyi, T. 2005, *MmSAI*, 76, 1004
Mezö, G., Baranyi, T., & Györi, L. 2005, *Solar Magnetic Phenomena*, 320, 247
Murtagh, F. & Heck, A. 1987, *Astrophysics and Space Science Library*, 131,