Mon. Not. R. Astron. Soc. **323**, 223–230 (2001)

Comparison of sunspot area data bases

T. Baranyi,[1★] L. Győri,[2] A. Ludmány[1] and H. E. Coffey[3]

[1]*Heliophysical Observatory of the Hungarian Academy of Sciences, Debrecen, PO Box 30, H-4010, Hungary*
[2]*Heliophysical Observatory, Gyula Observing Station, 5701 Gyula, PO Box 93, Hungary*
[3]*NOAA/NESDIS/NGDC E/GC2 Solar–Terrestrial Physics Division, 325 Broadway, Boulder, CO 80303, USA*

Accepted 2000 November 7. Received 2000 October; in original form 2000 August 24

ABSTRACT

Sunspot area measurements play an important role in the studies of sunspot groups and variations in solar irradiance. However, the measured areas may be burdened with systematic and random errors, which may affect the results in these fields. Mainly the total solar irradiance models can be improved by using more precise area data. In order to choose the most appropriate area data for a given study or create a homogeneous composite area data base, there is a need to compare the sunspot areas provided by different observatories. In this study we statistically investigated all the available corrected sunspot area data bases for the years 1986 and 1987. We find that the photographic data bases are in good agreement with each other but there are important systematic differences between the photographic and sunspot drawings data bases. We give the characteristic parameters for the systematic and random errors as well as the possible reasons for them.

Key words: methods: data analysis – sunspots.

1 INTRODUCTION

Sunspot areas are important parameters for lots of studies in various fields of solar physics: growth and decay of spots (Lustig & Wöhl 1995), emergence of fluxes in spots (Baranyi & Ludmány 1992), evolution of sunspot groups and interaction between them (van Driel-Gesztelyi et al. 1993), axial tilt and rotation rate of groups (Howard 1991, 1992), periodicities in solar activity (Oliver & Ballester 1995), fragmentation of flux tubes (Ludmány, Baranyi & Mező 1999) and irradiance variability (Fröhlich et al. 1991).

Because of its importance the area of sunspots is measured in several observatories, but after the discontinuation of Greenwich Photoheliographic Results there is no widely accepted standard for these data. However, while determining the sunspot area one has to cope with many difficulties that result in random and systematic errors (Győri 1998). These errors may affect the results of the study in which they are used. In the best case they only cause scatter in the related data without distortion of the main result, but it could be a decisive effect in some cases.

The precision of sunspot area measurements is very important for studies of irradiance (Fligge & Solanki 1997). The total solar irradiance (TSI) is the value of the integrated solar energy flux over the entire solar spectrum arriving at the top of the Earth's atmosphere at 1 au. The space-borne irradiance observations began in 1978, and since that time the measurements reveal that the TSI varies by about 0.1 per cent over the solar cycle and by about 0.2 per cent over time-scales of days (Fröhlich & Lean 1998). It has been shown that the variations on a time-scale of days to decades are mainly caused by changes in the magnetic features on the solar surface (Foukal & Lean 1988; Chapman et al. 1992). In recent TSI models the sunspots, faculae and magnetic active network explain a considerable amount of the variations, but a significant part remains unexplained after removing the effect of sunspots and smaller magnetic elements (e.g. Fröhlich & Lean 1998). The question is whether this remaining variability is caused by unknown features or by the uncertainties of the data used in calculating these models.

At present the second case seems to be valid. The irradiance models use the concept of Photometric Sunspot Index (PSI) to account for the influence of sunspots on TSI. The sunspot area data are involved in construction of PSI in explicit form and in implicit form by the contrast of the sunspots (Beck & Chapman 1993). Fröhlich, Pap & Hudson (1994) found that improvement of the sunspot area data set improves the irradiance model significantly. That is why we think that there is a need for a comprehensive study in which the area data bases provided by different observatories are compared, so that we can reveal the systematic differences between them and the rate of the random errors. In this way one can choose the most appropriate area data for calculating the PSI and one can improve the TSI models.

2 OBSERVATIONAL DATA

The Debrecen Photoheliographic Data (DPD) catalogue (Győri et al. 1996, 1998) contains the whole area as well as the umbral area of the whole group and each spot in it. With this level of detail this catalogue is reckoned to be unique. The Debrecen data

★ E-mail: baranyi@tigris.klte.hu

are measured on daily white-light full-disc photographic plates as it was carried out in Greenwich. At this time the DPD is available for the years 1986 and 1987 in complete form. On the basis of these recently published catalogues we can compare the corrected (for foreshortening) area measured at different observatories.

A similar data base is published in the Solnechnie Dannie (SD 1986, 1987) based on measurements of white-light photographic plates but they are measured with a different method. This SD catalogue contains the areas of the whole group as well as that of the largest spot of the group. These two data bases provide areas for each day because they gather plates from several cooperating observatories.

The third photographic data base is based on plates taken in Rome (Solar Phenomena 1986, 1987, 1988). It contains the area of the whole group and the sum of the umbral areas measured in the group. Some groups are omitted from the area measurements, only their position data are published. Its coverage is limited as a result of the weather conditions and instrumentation problems.

The other data bases are based on sunspot drawings, and we refer to them as graphic data bases by analogy with photographic data bases.

The Solar Optical Observing Network (SOON) consists of a worldwide network of solar observatories located so that 24-hour synoptic solar patrol can be maintained. Sunspot drawings on an 18 cm diameter projected image of the Sun are made daily at each site. The basic observatories are Boulder and the members of the network of the US Air Force (Holloman, Learmonth, Palehua, Ramey and San Vito). In addition, some other observatories take the responsibility of this patrol (e.g. Culgoora). The data are encoded and communicated to the regional warning centres routinely. These encoded data and the sunspot drawings are then sent to NOAA NGDC for archiving and inclusion in the monthly report Solar–Geophysical Data (1986, 1987). Because SOON operates in a real time mode, time is of the essence and detailed, extremely accurate measurements are not possible. The scaling of positions and areas are performed routinely by hand, using Stonyhurst disc overlays for both elements. The SOON sites do not generally report a sunspot area less than 10 millionths of the solar hemisphere and the published areas are also rounded to this precision. This observing procedure is followed at all of the sites, giving an internal consistency, but no further screening for outliers or errors is done. In spite of its disadvantages, the SOON data base is used in the present TSI models, since it becomes quickly and widely available and it is the only sunspot region catalogue that gives relatively complete (80 per cent) daily coverage.

The daily sunspot observations published in the Chinese Solar Geophysical Data are based on the visual data that mainly come from the Yunnan Observatory (CSGD 1986, 1987). When there are gaps in these observations the table is filled by observations of other cooperating Chinese observatories. The table is standardized after collecting all sunspot observations from different observatories. This catalogue contains the areas of the whole group as well as that of the largest spot of the group.

In the Catania Astrophysical Observatory the daily drawings of sunspot group were made at the Cooke refractor (15 cm/223 cm) on a 24.5 cm diameter projected image of the Sun (Solar Observations 1986, 1987). They observe under very good average weather conditions as there were no gaps in the data base for the studied two years.

The DPD data are available from http://fenyi.sci.klte.hu and the other data are available from http://www.ngdc.noaa.gov/stp.

3 COMPARISON OF DATA BASES

By using their position data we identified the sunspot groups for the statistical investigation. The groups close to the limb were omitted from the study. We selected the groups whose relative distance (distance from the centre of the solar disc divided by the disc radius) was smaller than 0.98 in the DPD. In this way we obtain eleven tables in which there were data pairs of sunspot areas measured in Debrecen and selected from a given data base on the same group on the same day. We made a linear regression analysis and a curve estimation for these data sets in the form: dependent $= a + b \times$ independent. The independent variable was the area (D) published in the DPD. The dependent variables were the areas published in the eleven other data bases: Solnechnie Dannie (*SD*), Rome (*Ro*), Boulder (*B*), Catania (*Ca*), Culgoora (*Cu*), Holloman (*H*), Learmonth (*L*), Palehua (*P*), Ramey (*Ra*), San Vito (*SV*) and Yunnan (*Y*).

3.1 Area of the whole group

Table 1 shows the eleven results for the area of the whole group, and the first column in Fig. 1 shows the plots of five of them.

One can see that the photographic *SD* and *Ro* data are larger with a relatively big constant but neglecting this fact there is no systematic difference between them as b is practically equal to 1. Concerning the graphic data there is need to correct them with only a small constant but the areas are smaller than that of DPD by quite a large percentage. From this point of view the best is Catania because its areas are smaller than the photographic data by only 12 per cent, but Culgoora areas are smaller with 41 per cent. It is also remarkable that there are large deviations from the linear curve mainly for *SD*. From this point of view the best data are the Yunnan and San Vito areas.

To some extent the deviations between the data of different observatories can be explained by the fact that they observe at different times in UT. During a given time interval the evolution of sunspot groups can cause real differences between the areas. The value of the standard error caused by the evolution depends only on the difference of the geographic longitudes of the observatory measured from Debrecen. We do not see such a kind of dependence of the standard error. Thus, this kind of standard error can not be larger than the values calculated for the remote Yunnan. This means that the other data bases are burdened with larger random errors than the Yunnan data base. (It may also be true for San Vito as the standard error is about that of Yunnan but it is close to Debrecen.) We studied the question of whether the

Table 1. Results of the linear regression for the area of the whole group.

Database	Number of cases	a	b	Standard error of the estimate
SD	894	34.60	1.025	68.43
Ro	279	10.45	0.995	66.25
Ca	875	5.48	0.877	57.95
Y	800	0.71	0.806	44.11
SV	572	3.77	0.778	43.19
Ra	784	1.00	0.762	53.92
H	841	3.18	0.731	52.12
P	729	2.99	0.695	47.14
B	612	8.25	0.680	66.91
L	882	6.79	0.629	54.88
Cu	510	5.00	0.594	64.77

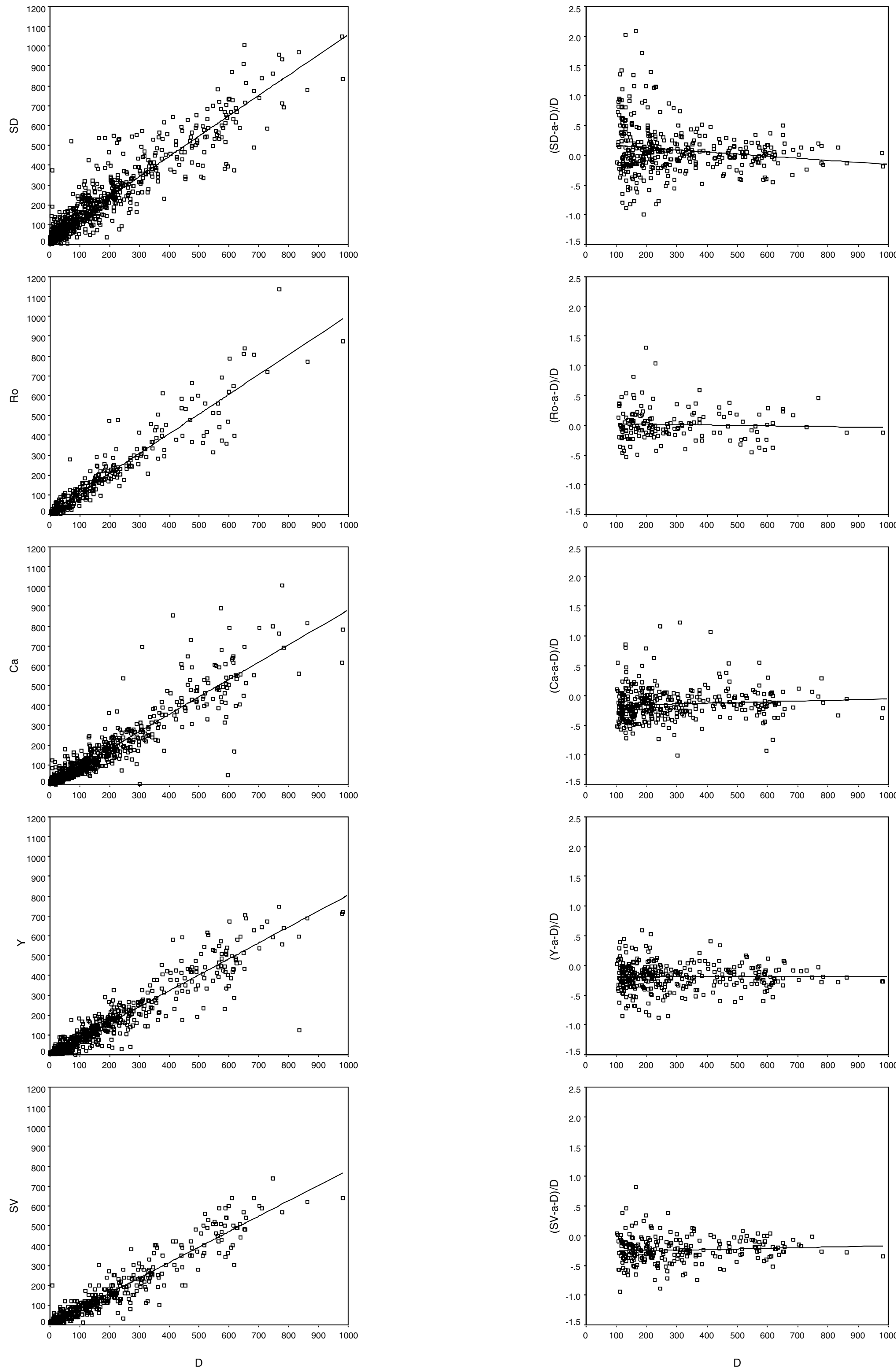

Figure 1. First column: Area of the whole sunspot group published in different data bases versus Debrecen area of sunspot groups. From top to bottom: Solnechnie Dannie (*SD*), Rome (*Ro*), Catania (*Ca*), Yunnan (*Y*) and San Vito (*SV*). Second column: Relative error of area of sunspot groups published in these data bases versus areas of sunspot groups measured in Debrecen.

systematic deviation depends on the area of the whole group, in other words, how would the $b-1$ value vary if we computed it for narrow ranges of area data. For this study we defined a new parameter named the relative error, which is the ratio of the difference (other observatories-Debrecen) to the Debrecen value, i.e., by taking into account the a constant difference we define the relative error by the following expression: (dependent-a-independent)/independent.

It is reasonable to omit the small groups from the examination of the relative error because, for instance, if someone measures 20 unit area for a group instead of the real 10 unit, it means only a small difference in the absolute value of the area (mainly in comparison with the areas of the large groups), but it means 100 per cent relative difference. To avoid this problem we studied only those cases where the Debrecen area $D>100$. Table 2 shows the results for the relative errors of the areas of the whole group, and the second column in Fig. 1 shows the plots of five of them.

These data indicate that, the *SD* data are larger than D with an average difference of 16 per cent at $D=100$, but the difference quickly decreases with the increase of D. Meanwhile the *Ca* data are smaller than D with the same difference at $D=100$, but the absolute value of the difference also decreases with increasing D. For the largest groups practically there is no difference between these data bases. Most of the data bases based on sunspot drawings show the similar increase of b with increasing D but they can not reach the zero value. However, Yunnan and Boulder show a negligible increase of b, and their areas are smaller than D with a constant percentage. We do not find any important systematic difference between *Ro* and D data.

Concerning the scatter of the relative errors for all data bases we conclude that it is quite large at about $D=200$ but it also decreases with increasing D. The standard error in Table 2 gives the average value. From this point of view Yunnan and San Vito show the best result again. Their average random difference from DPD is only ± 22 per cent.

In the previous paper (Baranyi et al. 1999) we obtained the following values by using the data of Kislovodsk plates from DPD and the data for the same day from *SD*: $a=33.17$, $b=1.015$. The value of a is almost the same as that for all D and *SD* data and we concluded that the constant difference between D and *SD* is systematic. If the relationship of two data sets is really linear then the constant should be zero. If it has a non-zero value then the connection of data is non-linear, which might result from differences in the methods of evaluation. (This method is thought to be responsible for the larger errors close to the limb.) Apart from this constant, almost the same areas are measured from the same plates in both data bases.

Table 2. Results of the linear regression for the relative error of the areas of the whole group.

Database	Number of cases	a	b	Standard error of the estimate
SD	392	0.194	−0.00035	0.406
Ro	159	0.030	−0.00006	0.272
Ca	362	−0.175	0.00011	0.273
Y	349	−0.188	−0.00001	0.221
SV	250	−0.273	0.00010	0.222
Ra	341	−0.290	0.00009	0.277
H	370	−0.347	0.00016	0.255
P	334	−0.364	0.00012	0.236
B	311	−0.335	0.00003	0.313
L	371	−0.436	0.00014	0.235
Cu	269	−0.495	0.00019	0.260

3.2 Area of the largest spot

The small spots and the fragmented amorphous penumbral parts usually are measured with quite a large uncertainty. These uncertainties contribute to the errors of the area of the whole group. However, it would be important to know the accuracy of the measurements of single spots. Therefore, we chose those cases where it seemed to be unambiguous that the same large and well-defined spot was measured as largest spot for DPD, the SD, and Yunnan.

The first column in Fig. 2 and the Table 3 show the result of a linear regression between the areas of the largest spots of DPD (*Ds*) and that of Solnechnie Dannie (*Ss*) and Yunnan (*Ys*). Comparing these parameters with that obtained in the previous subsection, we see some important differences. For the *Ss* data the a constant and the standard error decreased considerably but the b has practically the same value. This means that areas of well-defined individual spots are less burdened with errors but the *Ss* data are systematically larger with a constant of about 19 and a percentage of 5 per cent. For *Ys* the random errors are smaller than for *Ss* as found in case of the area of the whole group. However, there is a surprising result here: the *Ys* data are smaller than *Ds* by only 6 per cent, which is much less than the 19 per cent found in case of the whole group.

We studied again how the systematic error depends on the area. The second column in Fig. 2 and the Table 4 show the result for the relative error in case of $Ds>100$. Comparing these parameters with those in Table 2 we see that there is no significant difference between the results for *SD* and *Ss* data. This means that the relative error is similar in the case of individual spots and the whole groups, but there is a little bit larger (5 per cent) difference between the two data sets. Concerning the *Ys* data, the deviation (b) for the single spots is substantially smaller than for the whole spots. The most important thing is that the relative error depends on *Ds* in a similar way as we found with the *Ss* data but in opposite sense. This means that at $Ds=100$ the *Ss* data are larger by 14 per cent while *Ys* are smaller by 12 per cent but the relative error of the three observations is practically zero at about $Ds=500$.

By using only the areas of largest spots measured on Kislovodsk plates from DPD and from SD, we obtained the following values: $a=19.05$, $b=1.002$. The essentially unchanged a confirms that this non-linearity results from the method of measurement. Taking into account this constant there is no systematic difference (0.2 per cent) between *Ds* and *Ss* measured on the same plates. This means that the 5 per cent difference in Table 3 comes from real differences between the plates used for DPD and Solnechnie Dannie.

Summarizing the above results we conclude that for the well-defined spots the random errors are smaller than for the whole groups. However, there is a systematic difference: the areas of Solnechnie Dannie approach the DPD data from above while the *Ys* data approach the DPD data from underneath with the increasing area.

If the areas of the well-defined *Ys* spots approach the *Ds* with increasing area whereas the relative error is constant for the whole group, this can be interpreted in such a way that some smaller spots and penumbral parts might have been omitted from the measurement of the whole group. For any area of sunspot group

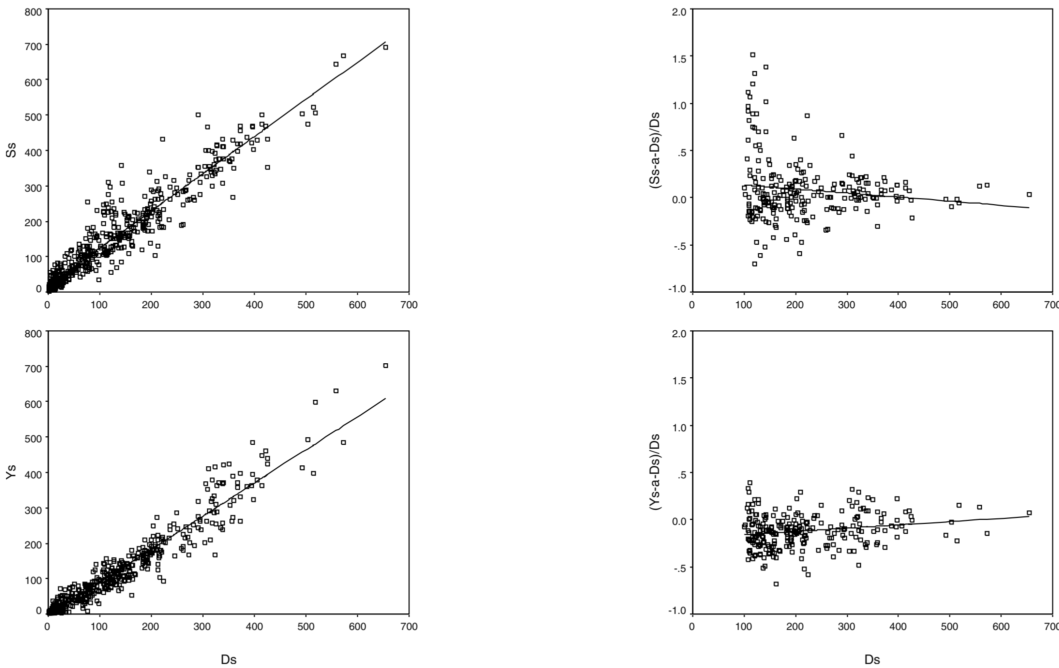

Figure 2. First column: Area of the largest spots published in the Solnechnie Dannie (*Ss*) and in Yunnan (*Ys*) versus area of the largest spots measured in Debrecen(*Ds*). Second column: Relative error of area of largest sunspots published in these data bases versus area of largest spots measured in Debrecen.

Table 3. Results of the linear regression for the area of the largest spots.

Database	Number of cases	*a*	*b*	Standard error of the estimate
Ss	519	18.88	1.051	40.18
Ys	519	−4.58	0.937	30.31

Table 4. Results of the linear regression for the relative error of the area of the largest spots.

Database	Number of cases	*a*	*b*	Standard error of the estimate
Ss	236	0.177	−0.00043	0.330
Ys	236	−0.150	0.00026	0.178

about 19 per cent of the area is neglected. The situation is similar at Boulder but 33 per cent of the whole area is omitted here. For the other graphic data bases the relative error decreases with increasing area, so they neglect a decreasing part of the whole area but only Catania can reach the zero value.

3.3 Area of the umbra

Rome publishes the sums of areas of umbrae measured in the group (*Ru*), so we can compare them with the similar DPD data (*Du*). The results of the linear regression for 278 umbra areas are: $a = 6.13$, $b = 1.15$ and the standard error is 20.51. Concerning the relative error for 55 cases when $Du > 50$: $a = 0.487$, $b = -0.033$ and the standard error is 0.474. Fig. 3 shows the related plots.

Comparing these values with the data in Table 1 and 2, we see that the constant is about the same for *Ro* and *Ru*, and it is likely to be the result of the measuring method. However, the *b* is much larger for *Ru* than for *Ro*. The relative error for *Ru* is 32 per cent at $Du = 50$ but it decreases quickly with increasing *Du*. Its average is 15 per cent which is computed for the whole range of *Ru*. This behaviour is very similar to what we see for *Ss* but its extent is much larger for *Ru*. This may refer to the same effect of low gamma which causes the *Ss* areas to be larger. For *Ro* data we detect this effect to a less degree. Therefore, we think that neglect of some part of the whole groups compensate this effect to some extent. Thus, the Rome areas for the whole group show only a little systematic deviation but the umbra areas are much larger than DPD umbra areas.

4 CAUSES OF THE DEVIATIONS

The differences between the measured areas can be explained with several effects. Repeated measurements of a given observer show a few percentage random differences, as one can find the same contours quite precisely. Comparing two or more observers we find smaller systematic and larger random errors (Gerlei 1987). The reason for the random errors is that in the case of spots with faint contours or with different brightness on their different parts, the personal bias can play an important role. The personal bias also causes systematic deviation as somebody tends to find the contour closer to the centre of the spot while the other observer tends to find it at a larger distance. Area measurements of the same spot at the same time from different observatories may differ because of the limited seeing, which smears sunspot images and reduce contrast (Fligge & Solanki 1997). The different data reduction methods also cause large deviations between the areas (Pettauer & Brandt 1997). Influence of their combined effect was shown by studying the robustness of various methods for measuring sunspot areas under different seeing conditions (Steinegger, Bonet & Vázquez 1997). The difference in the time of observation also contributes to the random error as mentioned earlier.

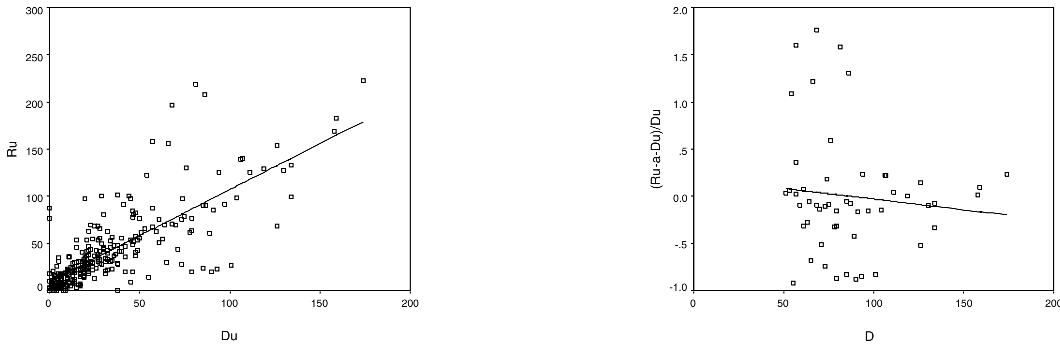

Figure 3. Umbral area of the group published in Rome and its relative error versus umbral area measured in Debrecen.

However, on longer time-scales when data are gathered from several observers in the case of different seeing conditions, there are only two things that mainly influence the systematic differences in addition to the average seeing conditions: these are the observing techniques and the measuring techniques.

As the result is the same for the same plates in DPD and SD, we have to seek for the cause of the differences in the fact that the types of plates (emulsions) used by DPD and by SD are different. From this point of view the most important parameter is the gamma of the film. The film response to the electromagnetic radiation is characterized with the H&D (calibration) curve showing the relationship between exposure and density, named after Hurter and Driffield, pioneers in the study of sensitometry. The H&D curve gives the optical density, or the logarithm of the film darkness, as a function of the logarithm of the relative exposure. In the optimal case the whole heliogram falls on the linear portion of this curve between the over-exposed and the under-exposed exposure range. The gamma is the slope of the linear portion of the curve. A film with steep slope shows high contrast over this linear portion: a small difference in the intensity can cause a large difference in the density of the film. If the gamma is low, features close in space and brightness may remain unresolved because the difference in density is negligible. At low gamma next to the well-exposed photosphere, the dark umbra and some part of the penumbra may be under-exposed, and this large under-exposed part is measured as umbra. If the gamma is very high, the intensity interval may be wider than the linear portion of the curve, and some part of the group may remain unresolved again.

The variation of the intensity across a spot to the umbra in the ideal case is the following (see fig. 1 in Beck & Chapman 1993): approaching the spot there is a narrow zone where the intensity decreases from the level of photosphere to the level of penumbra and coming through the plateau of the penumbra there is a similar zone where the intensity decreases from the level of penumbra to that of umbra. The points of maximum slopes in these zones are usually defined as the boundaries between photosphere and penumbra as well as penumbra and umbra.

Taking a photo we make a non-linear intensity transformation by the H&D curve. Because of this transformation, the points of maximum slope can be closer to the penumbra and umbra than they are in fact. The higher the gamma of the film is, the closer these points are. The gamma of the film used in Debrecen is about 6 depending on the exposition and development. We found that this results in areas that are smaller by about 5–10 per cent than those that would have been obtained without the non-linear transformation induced by the film. The thickness of the zone between the photosphere and penumbra does not change linearly with the extension of the spot. So the larger the spot is, the smaller the relative error between the areas measured on films of different gamma. This explains the decrease of the relative error with increasing area for data measured on films of lower gamma. For the films of higher gamma the areas are smaller than DPD data but they approach these data with increasing area.

5 DISCUSSION

On the basis of the observing methods we divided the area measurements into two subsets. One of them consists of the data bases in which the areas are measured on photographic plates, the other one comprises the areas measured on sunspot drawings. There are important systematic differences between the data of the two subsets, but smaller deviations also can be found between the data measured by the same method in addition to the random errors. Considering the causes of deviations we can determine the most probable reasons of the obtained results.

5.1 Data bases based on photographic plates

The photographic data bases studied are DPD, Solnechnie Dannie and Rome. In an indirect way we draw into the study the Greenwich Photoheliographic Results (GPR).

The gamma of the film is higher for Debrecen than for Solnechnie Dannie. One obtains larger area values on the film of lower gamma but the relative error decreases with increasing area. This explains that the individual *Ss* areas are larger with an average of 5 per cent than the *Ds* data and the exact value of the difference depends on the area. For the area of the whole group we found smaller (2.5 per cent) deviation. This may be explained because on a film of low gamma the small spots or penumbral parts may remain indistinguishable from the photosphere. This neglect causes that the areas of the whole group are closer to the DPD data than the areas of individual spots. The found non-linearity is shown to be the result of the measuring method of SD.

For the Rome data the results are similar. The constant deviation also seems to be the result of the measuring method. Considering the results for the areas of umbrae, we may suppose that the gamma of the used film is somewhat lower in Rome than in Debrecen. The umbral data are larger by 15 per cent on average but the precise value of this difference quickly decreases with increasing area. In contrary, the areas of the whole groups are almost the same as in DPD. The causes of this difference may be the unresolved parts of the group as was supposed for SD.

There is no overlap between GPR and DPD at present so we cannot compare them directly. However, there is a comparison between GPR and Debrecen Photoheliographic Results (Gerlei 1987). As the method of measurement is the same for DPR and DPD, it is assumed that these results are also valid for DPD. Thus, we can say that the GPR areas are larger by about 8 per cent than those of DPD because of the different method of measurement.

Summarizing the above we conclude that the photographic data bases provide nearly the same areas if the random errors and the constant deviations are neglected. Apart from the systematic differences caused by the photographic plates used or the method of measurement, the areas are in good agreement with each other.

This finding is in contradiction with the result of Fligge & Solanki (1997). They showed that the Rome data are smaller than GPR areas by about 20 per cent, but this contradiction may be resolved easily. They plotted the daily sums of Rome and GPR areas versus Zürich relative number. However, some groups were omitted from the area measurements in Rome. Therefore, the daily sums may be smaller than actual values, while for specific groups the area measurements give the appropriate results.

5.2 Data bases based on sunspot drawings

All the data bases measured on sunspot drawings provide systematically smaller areas by several percentage than the photographic data. The parameters characterizing the differences span quite a wide range. The exact value of the percentage depends on the average seeing conditions of the observatory as well as the parameters of the observing telescope. However, there is no constant difference between the graphic data and DPD, which means that the connection between these data bases is linear.

The Yunnan areas of the whole spots are smaller than that of DPD by 19 per cent uniformly. The deviation for the single spots is substantially smaller, and the relative error depends on the area in an opposite way as we found for the SD data. The area of largest spot of SD is larger while that of Yunnan is smaller than DPD area by almost the same percentage. Their difference decreases with increasing area so the relative error of the three observations is practically zero for the large spots. Therefore, we conclude that the root of the deviation of Yunnan data is the same as that of SD data. This means that making drawings can be modelled with usage of the photographic plate of the higher gamma: position of the point of the maximum slope is practically at the edge of the penumbra and the part of the group with brightness close to the photosphere can remain unresolved. This explains the different results for the individual spots and for the whole groups. The larger the group, the larger the omitted parts can be. Thus, the relative error for the whole group does not depend or only slightly depends on the whole area.

In Boulder 32 per cent of the area of each group is omitted from the measurements. In the other data bases, the percentage of the neglected area slightly decreases with increasing whole area. In Catania the observers measure only about 12 per cent smaller areas than in Debrecen. San Vito, Ramey, Holloman and Palehua provide almost the same areas. Learmonth and Culgoora differ form DPD with the largest values. The range of the random error also varies a little bit from one station to the other.

We think that the cause of the systematic differences between the photographic and graphic groups of data is likely to be hidden in the different conditions of measurements. In the case of measuring photographic plates as well as making sunspot drawings, the main measuring ‘facilities’ are the human eye and brain, but the circumstances for these ‘facilities’ to determine the borders of spots are basically different. It is well known that the viewer’s eye (and brain) estimates the brightness on a given place of the retina on the base of its deviation from the mean of the intensity of the surrounding area. This means that we make relative measurements of brightness. Comparing the usage of films with making drawings, we find several differences which can influence our perception. The negative film gives a gray background, so the average brightness is smaller than that of a projected image, and the rods can play a larger role in perception than in the case of drawing. On a film the interval between the intensity of brightest and the darkest points is less because it suffers a logarithmic transformation by the H&D curve. By measuring a film we can change the illumination to the optimal level by changing the voltage of the lamp. These differences may contribute to the result that the spots seem to be smaller when we make a drawing compared to the case when we look at a photo. Concerning the area differences of the whole group, by making a drawing we may neglect a fraction of the group. A similar effect can be experienced if we look at positive prints of different hardness. Brighter details may be missed on the harder photographic paper. Thus, the results for graphic data bases can be modelled by use of extremely hard films.

6 CONCLUSION

On the basis of this study one can choose the best data base for a given study or one can estimate the errors which get into the results of the study by the errors of the area measurements. We conclude that from the data bases with complete coverage the DPD is burdened with the smallest systematic and random errors. Unfortunately, the DPD is available only for a few years at present, though major efforts are devoted to achieve a full coverage. For studies on longer time-scales we can recommend the usage of Yunnan data base if umbral areas are not required. The CSGD provides area data for each day with the smallest random error. The systematic difference can be eliminated by multiplying the data with 1.24. The area data in DPD are smaller by about 5–10 per cent than the actual value because of the effect of hard film. Therefore, further increase by 5–10 per cent is reasonable mainly for studies of irradiance. By using the data in Tables 1 and 2 a composite data base can be created, which can have the largest daily coverage. By eliminating the systematic differences and comparing, screening and averaging the random errors the best sunspot area data base can be produced on the bases of these results.

ACKNOWLEDGMENTS

The work was supported by the grants of Hungarian National Foundation for Scientific Research Nos. OTKA T025640 and T026165.

REFERENCES

Baranyi T., Ludmány A., 1992, Sol. Phys., 139, 247
Baranyi T., Ludmány A., Győri L., Coffey H. E., 1999, in Proc. 9th European Meeting Sol. Phys., Magnetic Fields and Solar Processes, ESA SP-448
Beck J. G., Chapman G. A., 1993, Sol. Phys., 146, 49
Chapman G. A., Herzog A. D., Lawrence J. K., Walton S. R., Hudson H. S., Fisher B. M., 1992, J. Geophys. Res., 97, 8211

CSGD, 1986, Chinese Solar-Geophysical Data Monthly Reports. Academia Sinica, Beijing
CSGD, 1987, Chinese Solar-Geophysical Data Monthly Reports. Academia Sinica, Beijing
Fligge M., Solanki S. K., 1997, Sol. Phys., 173, 427
Foukal P. V., Lean J., 1988, ApJ, 328, 347
Fröhlich C., Lean J., 1998, in Deubner F. L., eds, Proc. IAU Symp. 185, Total solar irradiance variations: the construction of a composite and its comparison with models. Kluwer, Dordrecht, p. 89
Fröhlich C., Foukal P. V., Hickey J. R., Wilson R. C., 1991, in Sonett C. P., Giampapa M. S., Matthews M. S., eds, Sun in Time. Univ. Arizona Press, Tucson, AZ, p. 11
Fröhlich C., Pap J. M., Hudson H. S., 1994, Sol. Phys., 152, 111
Gerlei O., 1987, Publ. Debrecen Obs. Heliogr. Ser. 1, App. 2,
Győri L., 1998, Sol. Phys., 180, 109
Győri L., Baranyi T., Csepura G., Gerlei O., Ludmány A., 1996, Publ. Debrecen Obs. Heliogr. Ser., 10, 1
Győri L., Baranyi T., Csepura G., Gerlei O., Ludmány A., 1998, Publ. Debrecen Obs. Heliogr. Ser., 11, 1
Howard R. F., 1991, Sol. Phys., 136, 251
Howard R. F., 1992, Sol. Phys., 142, 47
Ludmány A., Baranyi T., Mező G., 1999, JOSO Ann. Rep., 1998, 99
Lustig G., Wöhl H., 1995, Sol. Phys., 157, 389
Oliver R., Ballester L. J., 1995, Sol. Phys., 156, 145
Pettauer T., Brandt P. N., 1997, Sol. Phys., 175, 197
Solar–Geophysical Data, 1986, No. 499–510, NOAA, Boulder
Solar–Geophysical Data, 1987, No. 511–522, NOAA, Boulder
Solar Observations made at Catania Astrophysical Observatory, 1986, Osservatorio Astronomico di Catania
Solar Observations made at Catania Astrophysical Observatory, 1987, Osservatorio Astronomico di Catania
Solar Phenomena, 1986, No. 326-340, Osservatorio Astronomico di Roma
Solar Phenomena, 1987, No. 341-354, Osservatorio Astronomico di Roma
Solar Phenomena, 1988, No. 355-366, Osservatorio Astronomico di Roma
Solnechnie Dannie Bulletin, 1986, Monthly Reports, Main Astrophys. Obs. Acad. USSR, Nauka, Leningrad
Solnechnie Dannie Bulletin, 1987, Monthly Reports, Main Astrophys. Obs. Acad. USSR, Nauka, Leningrad
Steinegger M., Bonet J. A., Vázquez M., 1997, Sol. Phys., 171, 303
van Driel-Gesztelyi L., Csepura G., Nagy I., Gerlei O., Schmieder B., Rayrole J., Demoulin P., 1993, Sol. Phys., 145, 77

This paper has been typeset from a TEX/LATEX file prepared by the author.