Numerical processing of sunspot images using the digitized Royal Greenwich Observatory Archive

Andrey Tlatov[1] and Vladimir Ershov[2]

[1]Kislovodsk Mountain Solar Station,Pulkovo observatory
[2]Mullard Space Science Laboratory, University College London (U.K.)

Outline

The Royal Greenwich Observatory (RGO) archive of solar photographic observations contains about 26000 plates obtained during the period of time from 1918 to 1972. This is a valuable series of solar data which can be used for analysing the influence of solar activity on Earth’s climate. The archive was digitised by Mullard Space Science Laboratory and converted into a series of FITS-format images with the resolution that corresponds to the sun's image radius having approximately 1600 16-bit pixel.

Special numerical algorithms have been developed for processing these images at the Kislovodsk Mountain Solar Station of Pulkovo Observatory (Russia). Our software includes also the manual screening tool for approving or rejecting the automatically localised sunspots. It is needed for removing photographic artefacts and marks (see Fig.1). We have localised about 140 thousand individual sunspots. The calculated areas of these sunspots were compared with the results of manual measurements of the same regions (www.ngdc.edu). The correlation between the automatic and manual results is quite high (R~0.93).

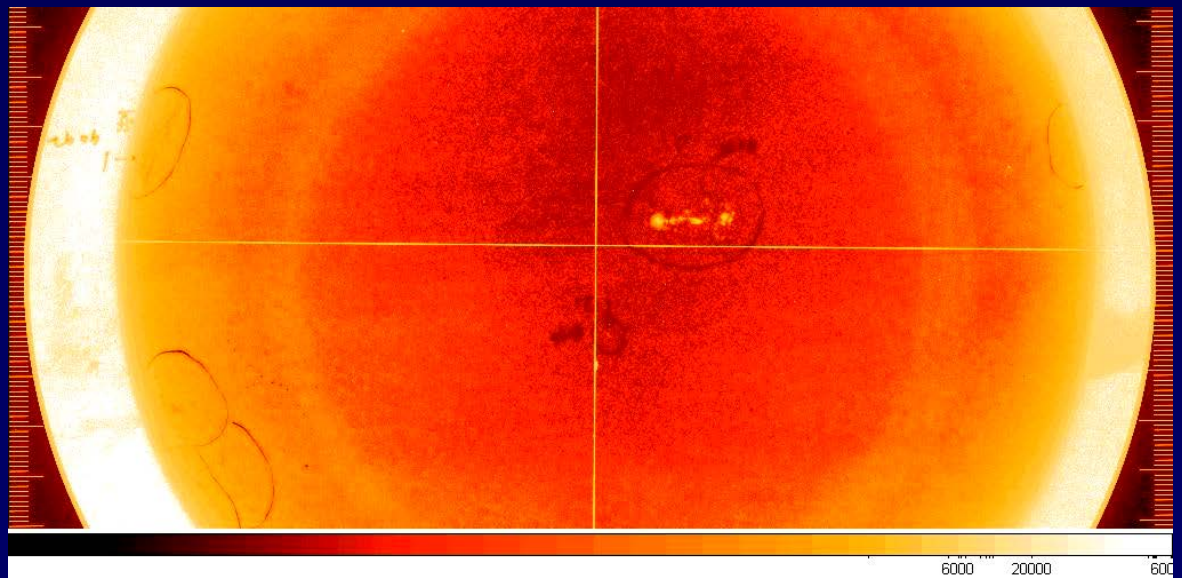


Fig.1. An sample from the digitised RGO archive showing artefacts and manual marks (Cape Town, 20th of January, 1920)

2. **Data preparation**

The preliminary processing stage of each photographic plate includes the identification of the edge of the solar disc and the overlaying of the heliographic coordinate grid that would correspond to the observation date and time.

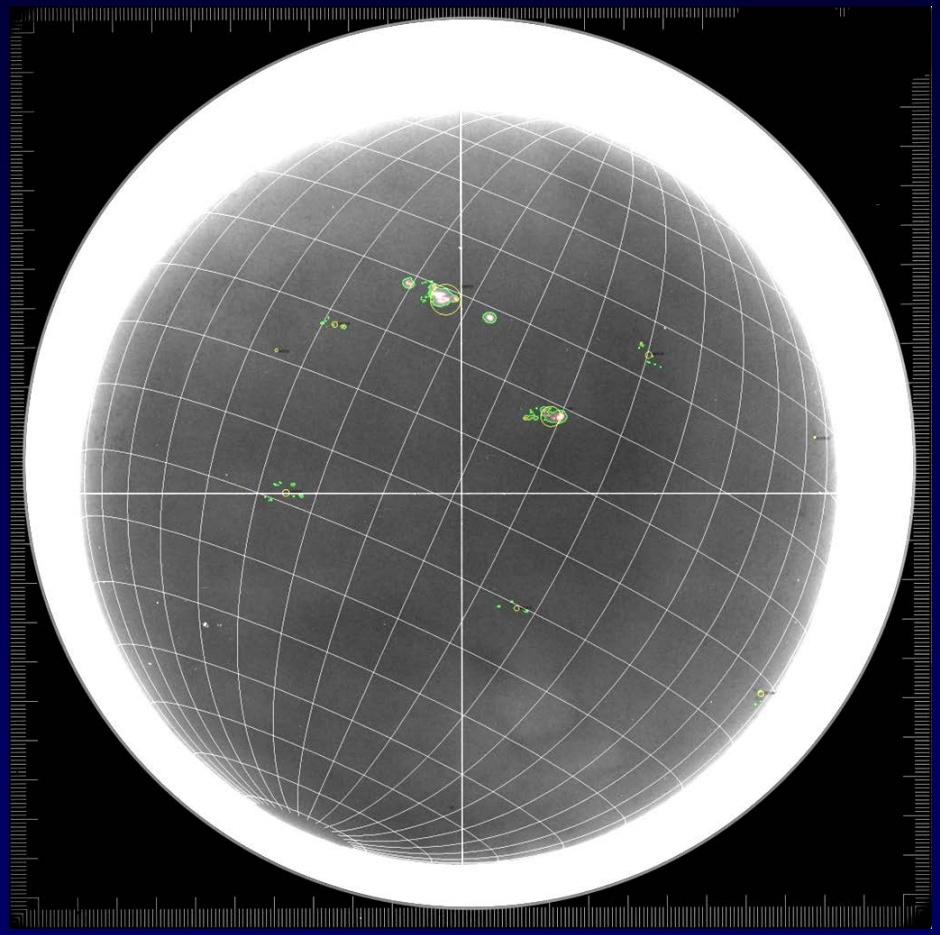


Fig.2. Sunspots identified on the heliogram dated 13.03.1918; they are compared with the sunspot group coordinates (shown as white circles) taken from http://solarscience.msfc.nasa.gov/greenwch.

3. Calculation of the sunspot parameters.

The second part of image processing consists of building a sequence of calibrated images. For our calibration we assumed that the center-to-limb variation of the quiet Sun (i.e., the variation calculated for the areas of the Sun without bright floccules) corresponds to a standard curve. The intensity local level of the quiet Sun (LQSL) is used for the elimination of non-uniformities across the solar disk, for image calibration and for searching elements of solar activity (active elements). For this purpose, we segmented the solar disc and searched for the maxima of the distribution pixel numbers for different intensities. This distribution corresponds to the level of quiet sun $I_{QS}(\theta,\varphi)$.

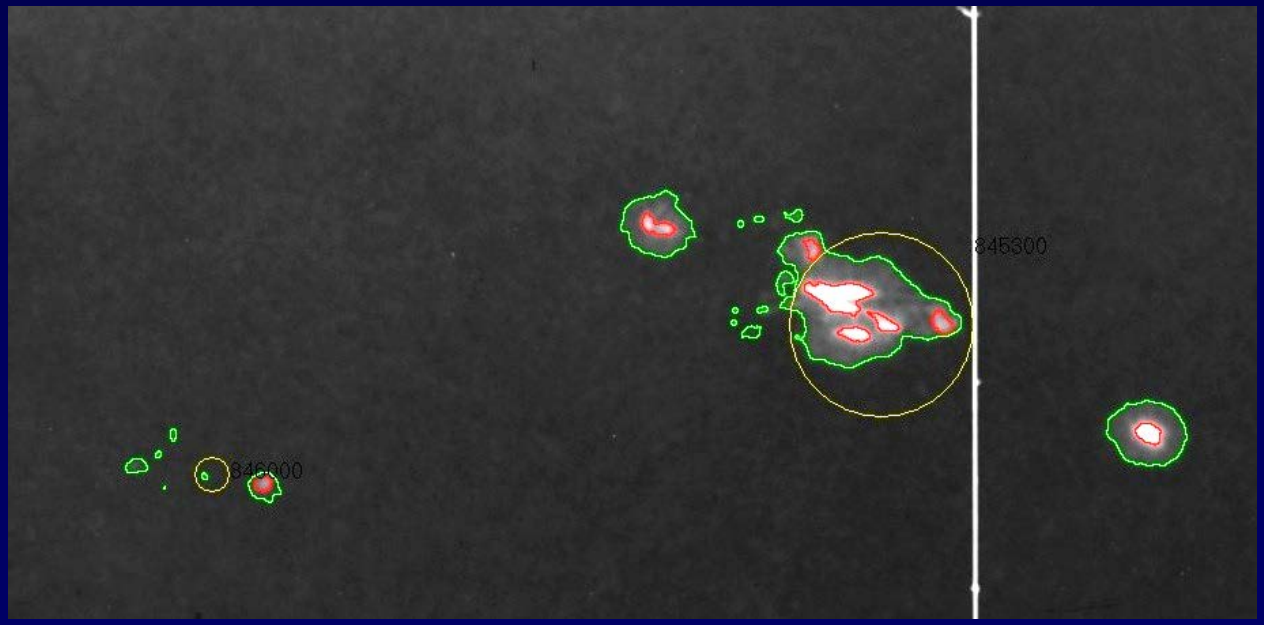


Fig.3. An example of the automatic sunspot detection using the image dated 13.03.1918. The green contours indicate the outer limits of the detected sunspots, the pink flags correspond to the darker cores. The yellow circle indicates the coordinates of the sunspot group measured manually (taken from the database).

4 Comparison of the automatic spot identification algorithm with manual measurements for the period of time from 1918 to 1972.

So far, the information about long-term solar activity was obtained by mainly using manual sunspot identification and also manual measurement of their parameters. It is difficult to verify the manual solar activity indices due to the lack of information about the geometry of the identified elements and not well-known procedures applied for these measurements. The use of automatic methods allows us keeping both quantitative and also vector information of the identified elements of solar activity. This also gives us an opportunity to determine some extra parameters related to these elements.

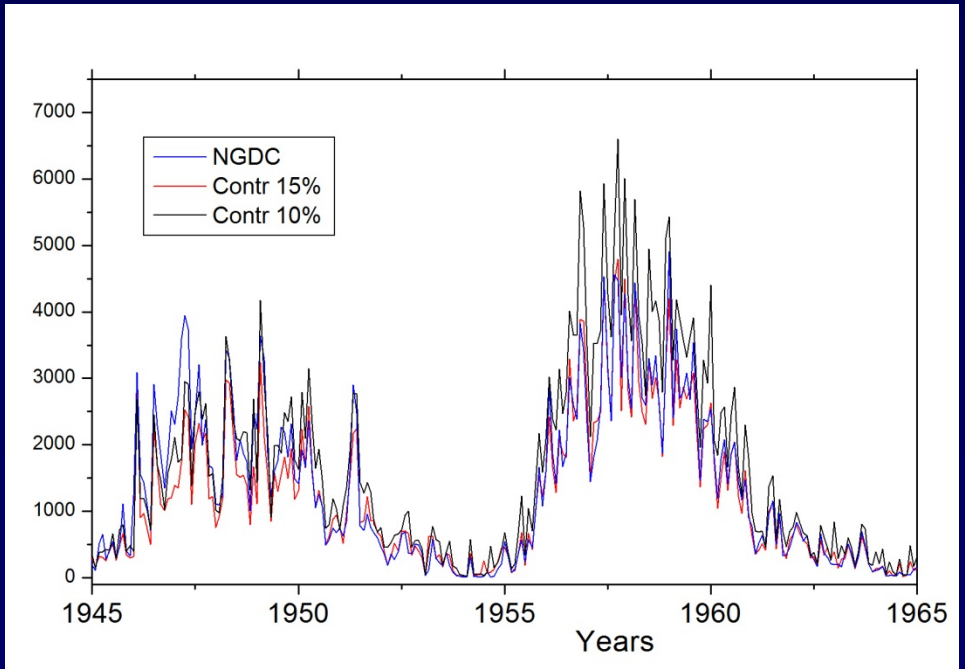


Fig.4. Comparison of the manually-measured sunspot areas (NGDC) with those obtained by using two different contrast thresholds (10% and 15%).

The comparison of the automatic sunspot identification procedure with the results of manual measurements taken from http://solarscience.msfc.nasa.gov/greenwch
In this case, the correlation coefficient between two data is R=0,93.

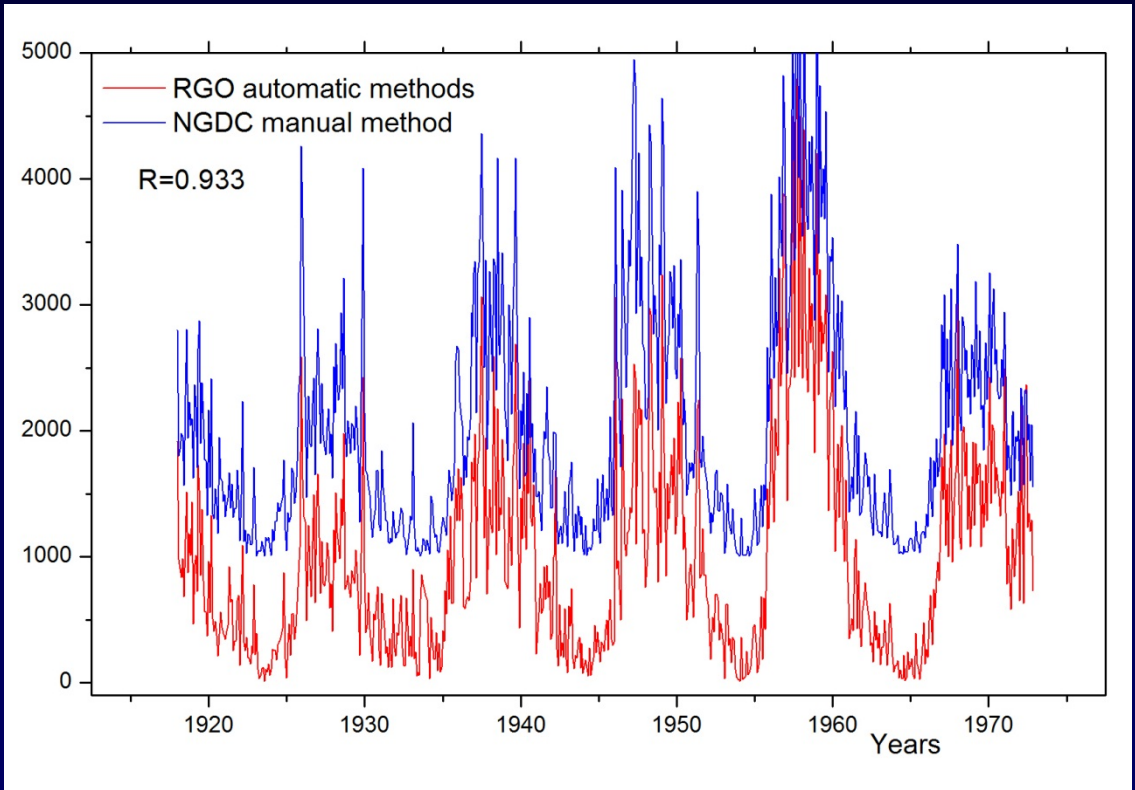


Fig.6. Comparison of sunspot areas obtained by using the automatic identification procedure and the sunspot areas measured manually. To avoid overlapping, the plot for the manual data is shifted by a constant value of 1000 μhm.

Different types of stored information.

1) Quantitative characteristics of the individual spots.

cape-19180113-110737_70.bp 1918/1/13 11:7 parameters: 0 15
R: 1716 xc: 1976 yc: 2029 P: 3.7 D: -4.4 L0: 227.6

q	fi	L	r/R	Smhm	Spx	dq	dfi	yrs	cntr	mn	mx	al	flux	lenht			Ngr
16.12	19.08	246.71	0.468	23.6	386.0	0.59	1.16	1927.0	0.29	1712	2316	-9.8	4.55e+04	0.4	246.1	247.3	8390
15.63	21.44	249.07	0.488	39.9	645.0	1.21	1.49	1991.5	0.33	1718	2714	-17.1	7.95e+04	0.8	248.3	249.8	8390
14.73	26.52	254.15	0.538	220.2	3435.0	2.78	2.46	2388.4	0.55	1749	5378	-26.7	5.26e+05	1.5	252.9	255.4	8390
13.30	32.66	260.29	0.601	102.6	1517.0	1.77	1.66	2282.4	0.40	1865	3849	-9.7	2.34e+05	1.0	259.5	261.1	8390
9.43	80.52	308.15	0.989	10.5	29.0	0.27	0.74	3759.3	0.18	3655	3971	-2.8	3.95e+04	0.5	307.8	308.5	8387
-14.94	3.30	230.93	0.191	26.4	479.0	0.80	0.93	1973.4	0.37	1661	2618	-7.4	5.20e+04	0.4	230.4	231.4	8392
-16.52	3.02	230.65	0.216	67.7	1223.0	1.73	1.79	1865.6	0.29	1662	2808	-24.6	1.26e+05	1.0	230.0	231.8	8392
-16.68	-1.19	226.44	0.213	18.8	340.0	0.55	0.95	1799.4	0.24	1673	2089	-4.5	3.38e+04	0.4	226.0	226.9	8392
-18.91	-5.18	222.45	0.264	105.3	1879.0	1.72	1.95	2027.0	0.39	1678	3133	-14.7	2.13e+05	0.9	221.4	223.4	8392
-19.30	-82.33	145.30	0.988	37.9	106.0	0.90	1.45	3380.1	0.18	3289	3543	10.7	1.28e+05	1.8	144.8	146.2	8398
-21.23	-82.54	145.09	0.989	16.0	44.0	0.47	0.82	3384.9	0.19	3275	3502	-3.9	5.42e+04	0.9	144.7	145.5	8398
-24.47	-82.62	145.01	0.989	10.6	29.0	0.30	0.72	3313.1	0.18	3207	3439	0.1	3.50e+04	0.6	144.6	145.4	8398
-25.65	-82.58	145.05	0.989	11.6	32.0	0.30	1.00	3322.0	0.19	3211	3488	-4.4	3.84e+04	0.7	144.6	145.6	8398
-21.91	82.48	310.11	0.989	15.9	44.0	0.44	0.94	3553.5	0.18	3468	3700	-1.7	5.64e+04	0.9	309.6	310.6	8396
-30.37	-82.60	145.03	0.989	12.3	34.0	0.31	0.81	3239.8	0.19	3147	3327	0.5	3.97e+04	0.6	144.6	145.5	8398
-23.62	82.48	310.11	0.989	17.2	48.0	0.57	0.88	3546.5	0.18	3453	3733	4.2	6.12e+04	0.9	309.7	310.5	8396
-30.47	82.59	310.22	0.989	20.5	57.0	0.56	1.07	3412.2	0.18	3311	3599	-0.9	7.01e+04	1.1	309.7	310.7	8396
-31.88	82.64	310.27	0.989	14.1	39.0	0.43	0.85	3384.9	0.18	3302	3535	3.3	4.78e+04	0.8	309.9	310.7	8396
-35.54	82.55	310.18	0.989	28.5	79.0	0.73	1.29	3346.6	0.20	3237	3518	-3.1	9.53e+04	1.3	309.4	310.7	8396

Total groups: 19 Area: 799.7 sig2_cn: 161.9 fc: 1425.3 sig2_ed: 190.5 feg: 2414.8 sig2_hl: 160.0 fhl: 2284.5

2) Vector of selected areas. These data can be used to convert, transfer to another image, to verify the data.

Some characteristics of the data

We have found the following relationship between the number of sunspot and their area: $S = -90(\pm 18) + 118(\pm 1) \cdot N$, with the correlation coefficient being $R = 0{,}94$.

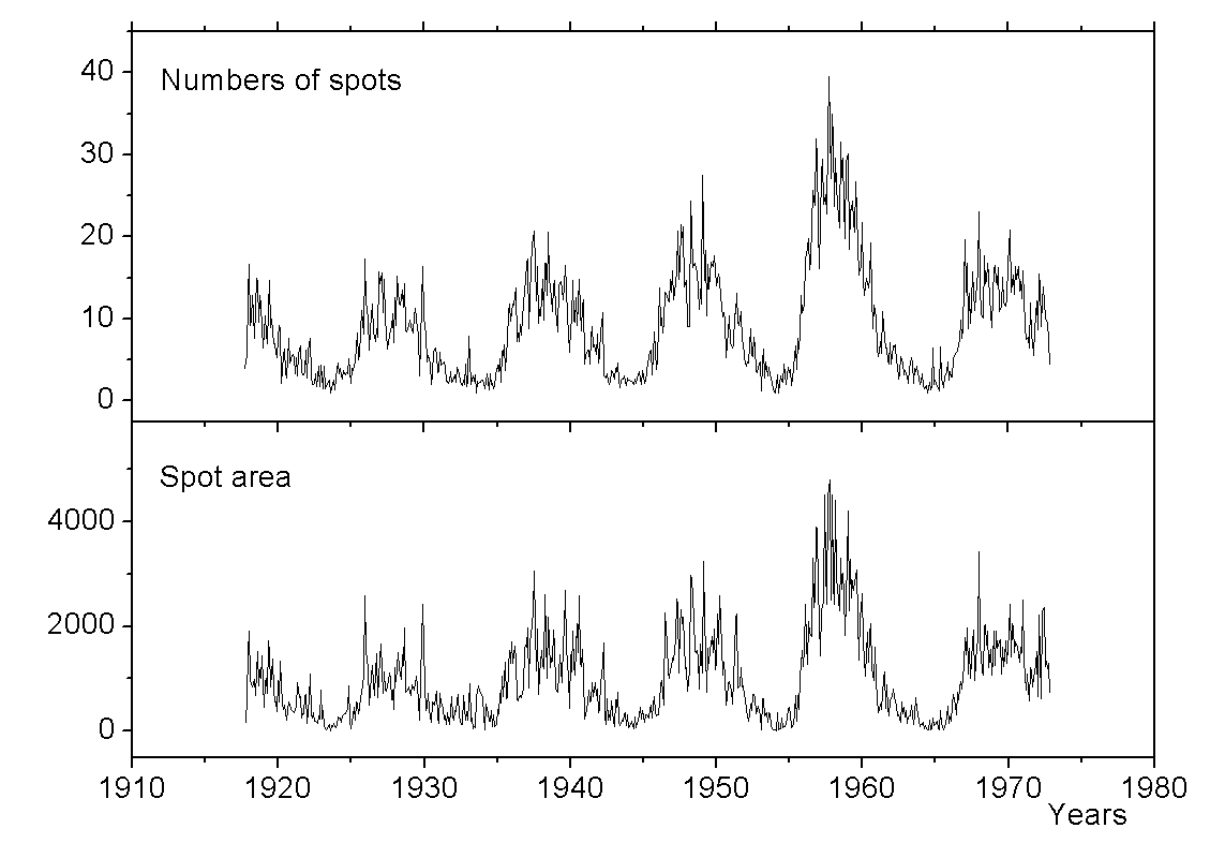
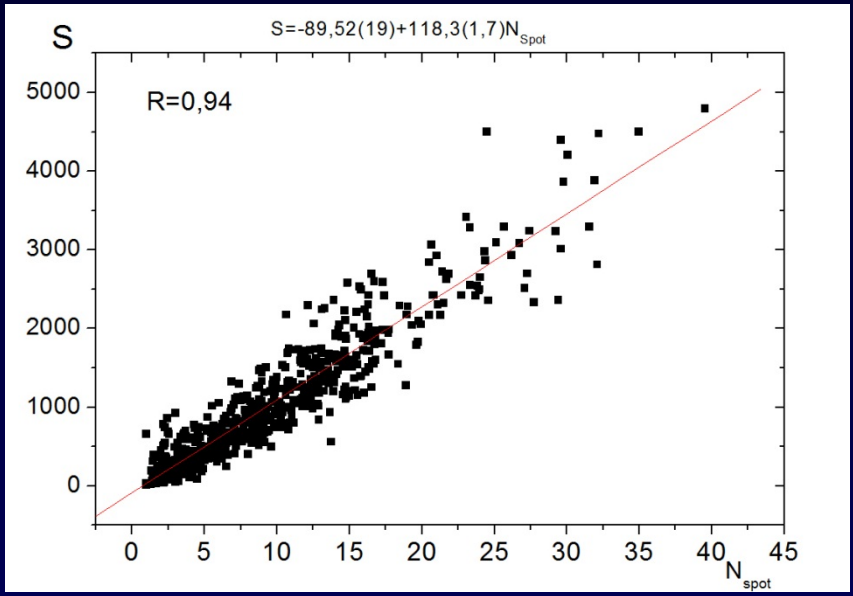
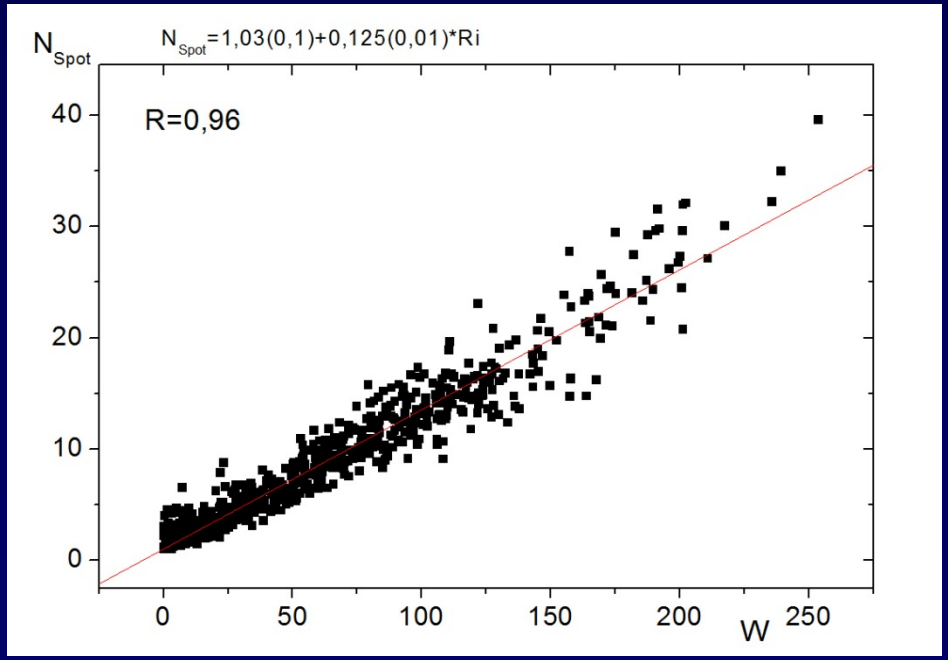


Fig.5. Upper panel: Monthly averages of the sunspot numbers calculated by using the automatic detection procedure; Lower panel: the monthly average sunspot area in 10^{-6} of the solar hemisphere area (μhm).

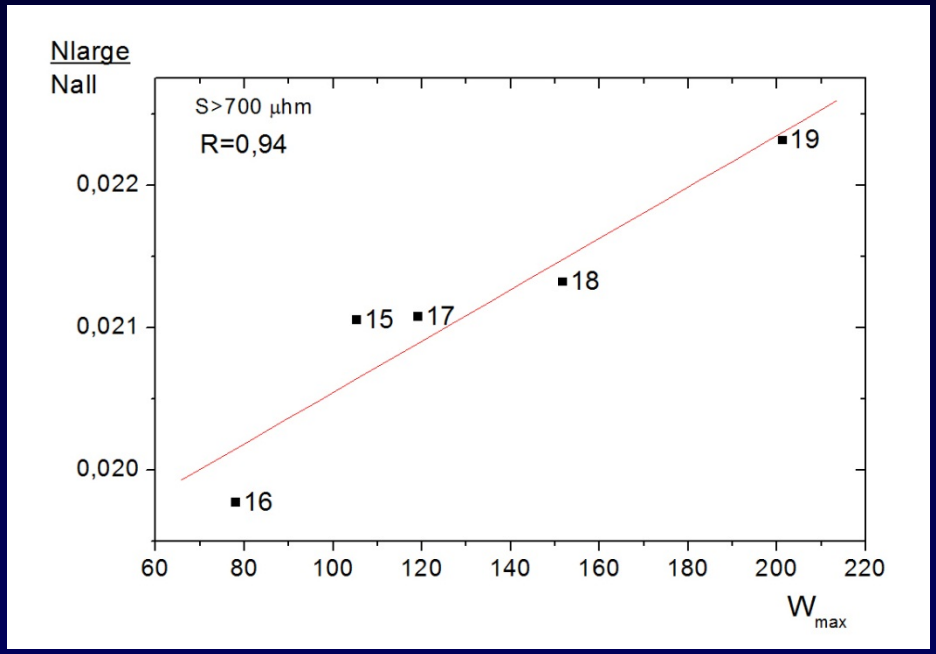
The total number of sunspots



The total number of spots can be used to check the sunspot index

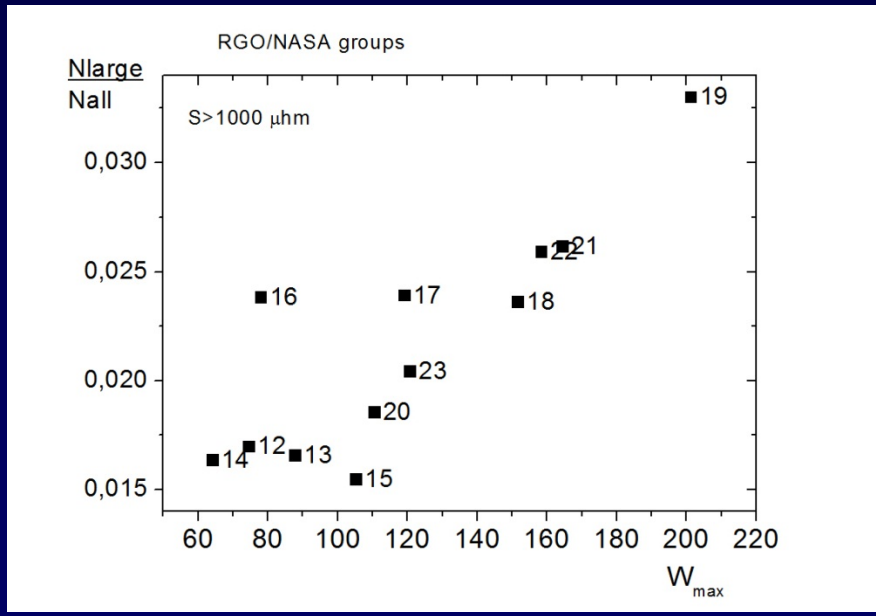


The relative number of large sunspot is linearly related to the amplitude of the cycle.



Individual sunspots S>700 μhm

Perhaps this is due to the lifetime of large sunspots



Sunspot groups S>1000 μhm

http://solarscience.msfc.nasa.gov/greenwch

Conclusions

We have presented our preliminary results of applying an automatic algorithm of sunspot identification to the historical series of photographic solar observation made by the Royal Greenwich Observatory. These results show good agreement with the manual measurements.

As a result of this work, a data base will be created containing large amount of information about individual sunspots, groups and their parameters.

We believe that these data as well as other image files (CaIIK from 1907, H-alpha from 1915 ….) must be available.

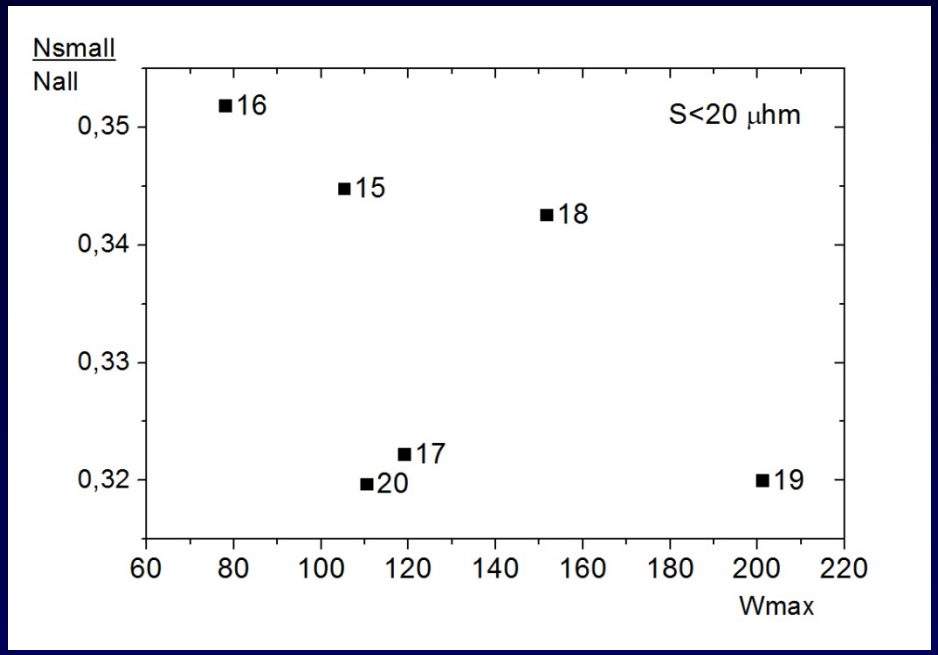
Questions

1)A large amount of image data (RGO ~1 TBt).

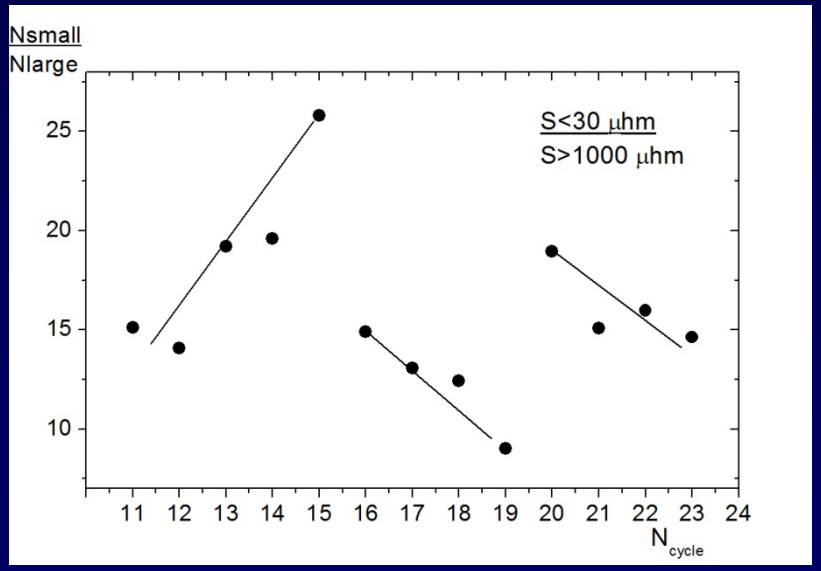
2)Should be able to recalculation the parameters to search for new indices such as fractal or change heliographic grid.

3) In what format we can provide this database: We believe that this should be the vector data format that can easily convert and edit.

The relative fraction of small sunspots varies from cycle to cycle



Individual sunspots S<20 μhm



Sunspot groups S<20 μhm

http://solarscience.msfc.nasa.gov/greenwch