COMPUTATION OF AMERICAN RELATIVE SUNSPOT NUMBERS

PETER O. TAYLOR
P.O. Box 1476
Boca Raton, FL 33429

Abstract

The current computation of American relative sunspot numbers is discussed, and brief comparisons between this index and Zurich-International sunspot numbers are drawn.

* * * * *

The American sunspot number, R_a, is an independent, relative index of daily and monthly sunspot activity. The importance of R_a is rooted in its reliability as an indicator of solar phenomena, and of certain solar-terrestrial relationships including geomagnetic variations and ionospheric effects. The related program of data-gathering and analysis has been coordinated by the American Association of Variable Star Observers through its Solar Division since the program's inception over forty years ago.

Herein is provided a brief description of present methods of data reduction and analysis.

Observer raw data are supplied as monthly sets of timed, daily values in the empirical form devised by R. Wolf in the nineteenth century:

$$R = 10g + f. \tag{1}$$

The factor "g" represents the number of sunspot groups visible to an observer, while "f" is equal to the total number of observed spots. The grouping scheme is the evolutionary classification system outlined by M. Waldmeier in 1938 and used at the Federal Observatory in Zurich for many years. Individual spots with an area of one twenty-fifth solar degree and larger are eligible for the counting process.

Before any raw sunspot data can be evaluated or reduced to actual sunspot numbers, it is first necessary to calculate two constants, k_i and w_i, for each potential contributor.

K_i is a scaling factor that compensates, in large part, for differences in equipment, average local seeing conditions, and personal judgment of the various observers. In the American program, k_i constants are computed from approximately one hundred suitable observations estimated during times when the monthly mean sunspot number equals or exceeds one hundred. New contributors to the program may be included at any time, however. Their data are necessarily restricted through an imposed lower statistical weight until monthly means have again returned to the target range, and k_i values re-computed.

The calculation equation from Shapley (1947) is equation (2):

$$a_i = \log k_i = \frac{1}{N} \left(\sum_{j=1}^{N} \log R_{sj} - \sum_{j=1}^{N} \log R_{ij} \right). \tag{2}$$

In (2), R_s represents a set of mean daily values used as the standard, in this case, the final American sunspot numbers for the computational period. R_i is the comparable set of data from a selected observer in the form dictated by (1), while N is the total number of data pairs.

To be useful for our purposes, each computer-generated k_i must fall within the range 0.5 to 1.5. However, for technical reasons, values skewed towards the lower end of the scale are preferred. An additional parameter requires that a contributor's k_i value vary not more than ten percent between semi-annual determinations. Estimates of R = 0 are not included in the computations.

The second constant, w_i, is a weighting factor, actually a measure of how well k_i corrects a given observer's unreduced data to the standard.

Equation (3), also from Shapley, is used in computation of w_i:

$$w_i = \frac{N - 1}{\sum_{j=i}^{N} (\log R_{sj} - \log R_{ij})^2 - N a_i^2}. \qquad (3)$$

Note that in equation (2), as Shapley has stated, the regression analysis is an example of the type described by Wald (1940) where both variables (in this case, members of sets R_s and R_i) may contain discrepancies. Again as Shapley points out, logarithms are chosen for (2) and (3) to produce a more homogeneous data set, one with the variance more equally distributed. This, and our assumption of uncorrelated errors, are basic requirements of the Wald method.

As data are received, each daily value, R, is multiplied by the contributor's k-factor, a process that minimizes the variance among all data for each day. The scaled data are then individually plotted with respect to day and six-hour Universal Time interval. Estimates of obviously poor quality that become apparent during the plotting process are not included in further computations.

Figure 1 is a refined example of such a plot. It should be noted that virtually all time intervals are well represented by observations. This provides a good indication of the scope of the AAVSO international sunspot observer network. In fact, contributors located on six continents contributed data that appear in the example.

Around the tenth of each month following observation, after a sufficient number (usually in excess of thirty) of reports have been received and initially processed, the computation of provisional sunspot numbers proceeds through application of the Shapley relation, equation (4):

$$R_a = \frac{\sum_{i=1}^{N} w_i k_i R_i}{\sum_{i=1}^{N} w_i}, \qquad (4)$$

for each day of the computational month.

Generally, N, the number of observers on a given day, will exceed eighteen or twenty. However, this number will depend to a considerable degree on the phase of the sunspot cycle, on prevailing world-wide weather conditions, and on observer confidence, especially during periods of minimal sunspot activity.

Final American sunspot numbers are reduced similarly, usually by the twenty-fifth of the month following observation, when virtually all reports have been received. The mean of the final values allows the calculation of another statistic, R_{sm}, the smoothed mean relative sunspot number. This number is computed with the aid of equation (5) from Waldmeier (1961):

$$R_{sm} = \frac{1}{24} (N_{i-6} + N_{i+6} + 2 (\sum_{-5}^{5} N_i)). \quad (5)$$

In equation (5), N is set equal to the final mean of the month under analysis, and other mean values are taken accordingly. Thus, R_{sm} lags approximately six months behind the most recent sunspot determinations.

Figure 2 is a graphic representation of values computed using equation (5) during the lifetime of the American program. Probably the results for sunspot cycles nineteen, twenty, and twenty-one (the present cycle) outweigh the earlier determinations, when some refinement of mathematical procedure was underway. Also depicted in Figure 2 is a plot of the smoothed 10.7 cm solar radio flux for most of the period. An excellent agreement between the two indices is demonstrated, especially after cycle eighteen.

Figure 3 is an intriguing comparison between American sunspot number monthly means (R_{am}) and those calculated at the Federal Observatory in Zurich (R_{Zm}) until 1981, and at the Royal Observatory of Belgium (R_{Im}) thereafter. Filled circles are an average of six monthly percentages obtained according to equation (6):

$$P_j = R_{am_j}/R_{Zm_j} \text{ or } R_{Im_j}; \quad j = 1 \text{ (January 1945) to } 474. \quad (6)$$

The increased amplitude of P_j prior to approximately cycle nineteen is obvious. It is interesting that R_{am} regularly exceeds R_{Zm} or R_{Im} only during times following sunspot maxima for perhaps a year or so. "Peaks" in the percentage time-line shown in Figure 3 appear to occur from six to twelve months after sunspot cycle maxima, while "valleys" coincide approximately with cycle minima. Perhaps this can be explained in part by differences in reduction methods between the two indices in the period before 1981. It is unclear why the effect should persist afterward (if it does continue, as it now appears) when reduction procedures are more nearly alike. No apparent correlation exists between heights of individual sunspot maxima and percentages approximating or greater than one hundred (i.e., times when R_{am} actually equals or exceeds R_{Zm} or R_{Im}), and probably none exists for the depths of corresponding minima either. The two indices do correlate extremely well, overall. Correlation coefficients of 0.993 are computed for data after 1966.5; of 0.994 for data after 1981.0; and, of 0.977 for all data since 1945.0. The average discrepancies for these time frames are 4.1%, 0.5%, and 5.6%, respectively.

The American sunspot number has proven valuable to the scientific community. The National Oceanic and Atmospheric Administration (NOAA) generously supplies the annual grant that funds a substantial portion of the work. The results are disseminated each month, either through the *AAVSO Solar Bulletin*, or in certain cases directly by telephone, to some 229 locations in 33 countries. These addresses include 57 domestic universities, libraries, and scientific organizations, 39 similar institutions located in 28 foreign countries, 63 contributors residing in 22 countries, and some 70 individuals. In addition, the index is reprinted in other scientific publications of amateur and professional interest, both here and abroad.

REFERENCES

Shapley, A. H. 1947, Publ. Astron. Soc. Pacific 61, 358.

Wald, A. 1940, Ann. Math. Stat. 11, 284.

Waldmeier, M. 1961, The Sunspot Activity in the Years 1610-1960, Zurich.

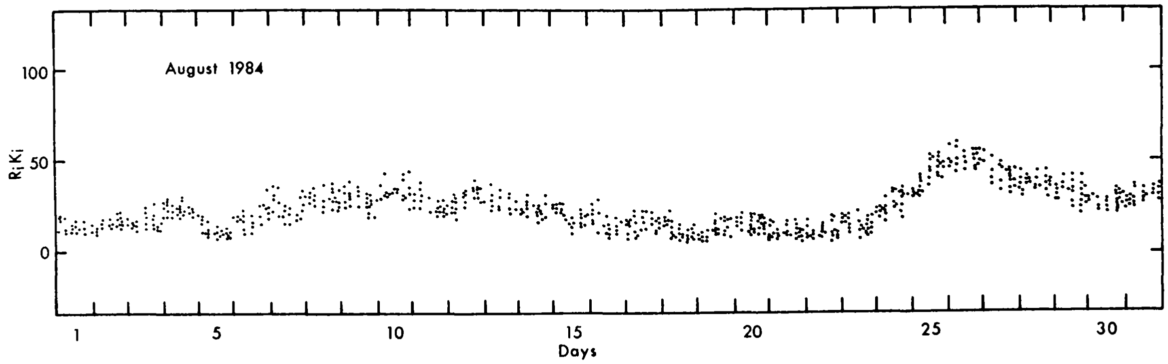

Figure 1. Each data point represents an individual observer's daily sunspot observation (in the form dictated by Equation (1)) after multiplication by the contributor's k-factor. Each estimate is plotted with respect to its six-hour time interval.

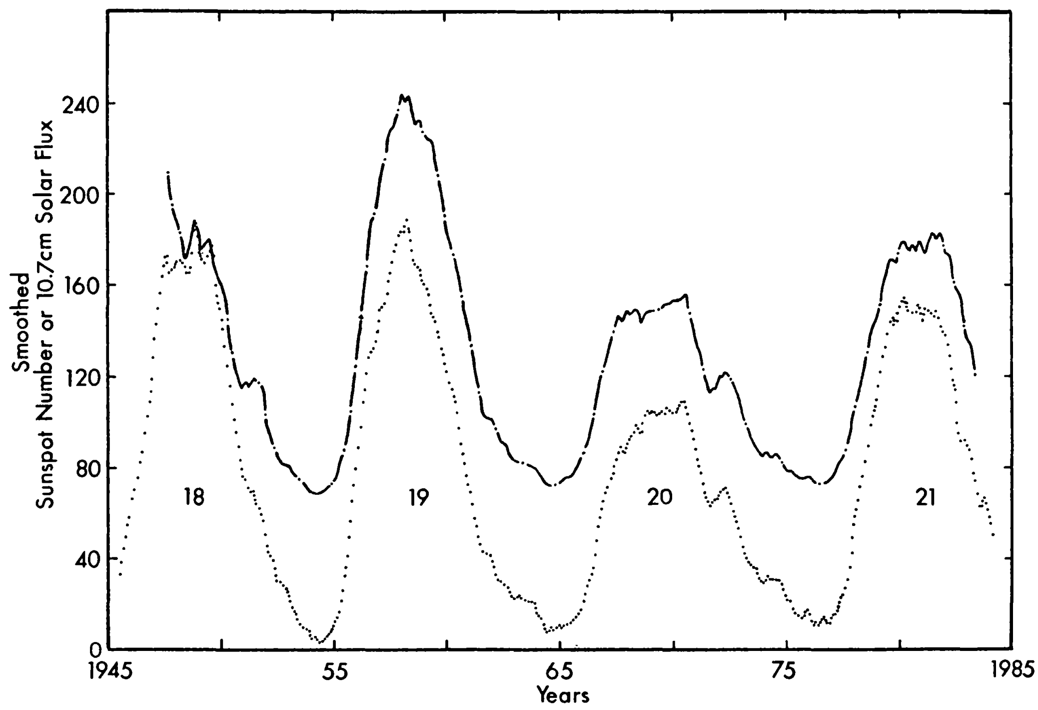

Figure 2. Individual data points represent values obtained using Equation (5). The upper curve represents the trend of 10.7 cm solar flux units (adjusted to 1 A.U.; Ottawa Series D) for most of the period. Sunspot cycles 18-21 are depicted.

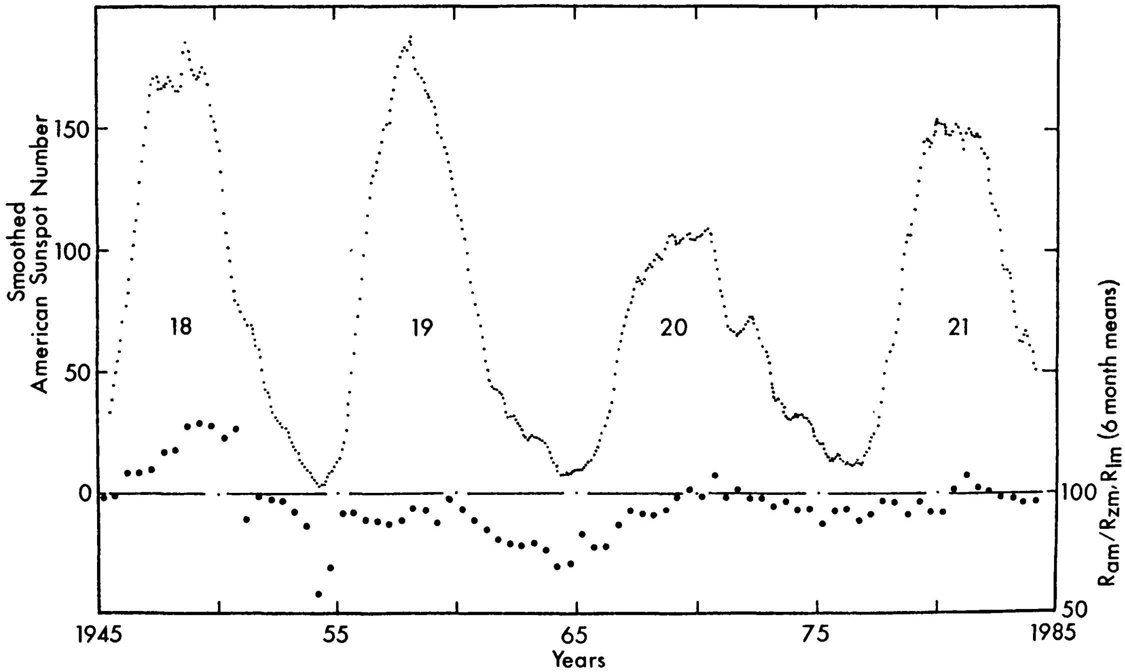

Figure 3. Small, individual data points represent values obtained using Equation (5). Filled circles are six-month means of percentage values computed with the aid of Equation (6). Sunspot cycles 18-21 are depicted.