AUTOMATIC INFLATION IN THE AAVSO SUNSPOT NUMBER

Bradley E. Schaefer
Physics Department, JWG 463
Yale University
New Haven, CT 06520-8121

Presented at the 85th AAVSO Annual Meeting, November 2, 1996

Abstract

The AAVSO sunspot number has a subtle mathematical flaw in its procedure which automatically inflates the reported numbers every time the K-coefficients are recalculated. This will result in a slow and spurious rise in apparent solar activity. This inflation is roughly 0.3% per recalculation, and the K-coefficients have been changed ~18 times since the founding of the AAVSO Sunspot Program in 1944. Thus, the current inflation is ~6%, and it will become exponentially worse with time. While the effect is now small enough not to be prominent, it causes the apparent level of solar activity to increase with time. This increase contributes to the confusion surrounding the issue of the detection of global warming due to greenhouse gases. Thus, the old AAVSO numbers must be retroactively corrected and procedures implemented that are free of inflation.

1. Introduction

Solar activity affects many aspects of life on Earth, ranging from activity/climate connections (e.g., the Maunder Minimum and the Little Ice Age) to failures of power grids. With the Space Age, the need for predicting and understanding our solar environment has become vital for many important purposes. For example, the observed global warming might be due to increased solar activity, man-made greenhouse gases, or some combination of the two. The disentangling of these two effects (and the consequent profound changes in public policy and long-term public health) requires a detailed and accurate knowledge of the long-term behavior of our Sun.

In recent times, precise and objective solar activity indicators based on radio and x-ray luminosities have been developed. While these are direct measures of the solar flux, they suffer from the shortness of the record. In contrast, sunspot numbers suffer from being an indirect measure of the solar flux, but have the great advantage of providing a record for almost four centuries (and for two millennia with naked eye reports). Thus, for studies of the long-term activity of the Sun (such as evaluating the greenhouse effect), the sunspot number is the only available and reliable index.

With so much at stake, we must examine the reliability of the sunspot numbers. The primary trouble is that all numbers explicitly rely on the observer's judgment as to what to call a spot (as opposed to a pore). In general, observers are widely scattered in their judgment, with typical variations of a factor of three. To avoid a bias based on the training or equipment of the individuals, all systems scale the counts to some standard and then average the scaled counts. But then there is the problem of how to maintain that standard over the decades. If the standard varies, then spurious increases or decreases in solar activity could lead to the implementation of incorrect public policy.

I have been analyzing the various sunspot numbers for the stability of standards. With regard to the AAVSO sunspot number, I have found a subtle mathematical flaw in the data reduction procedure which artificially inflates the published number with time. The consequences can be demonstrated from the published AAVSO procedure by means of a fairly simple mathematical derivation.

2. Inflation

The AAVSO procedure is reported in Taylor (1991). This procedure is identical to that described in Shapley (1949) and Taylor (1985). Essentially, the published number is a weighted and scaled linear average of all daily individual counts, where the scaling factor (the K-coefficient) for each observer is a logarithmic average of the previous year's final scale factor. I will denote the AAVSO sunspot number as R_l for the lth day of the current K evaluation period and r_j for the jth day of the previous K evaluation period. I will denote the observed count (i.e., the reported count of individual spots plus ten times the reported count of groups) to be C_{il} for the lth day of the current period and the ith observer, and c_{ij} for the jth day of the previous period and the ith observer. Similarly, let K_i be the utilized K-coefficient for the ith observer in the current K evaluation period, and k_i for the previous period. I will now examine the case where all observers are weighted equally (although I will later drop this assumption) for clarity of derivation. The equations presented in this section constitute a short derivation of inflation. Now, the AAVSO procedure has the daily published AAVSO number for the current period to be

$$R_l = \langle K_i C_{il} \rangle_i , \qquad (1)$$

where the brackets $<...>_i$ indicate a linear average of the available data over the given index. The K-coefficient used is derived with a logarithmic average:

$$K_i = 10^{<\log(r_j/c_{ij})>_j} . \qquad (2)$$

Note that the K-coefficients for the current period are evaluated by data from the previous period.

Now let us consider the identical observations as made by some perfect observer. This could be looked at as the result of some uniform and automated and sophisticated program operating off space telescope images. Such a perfect observer would report some true value for the sunspot number in the current period (P_l) and the previous period (ρ_j). Similarly, there will be some true K-coefficient for both periods for all AAVSO observers, which will be scattered around some average value for each observer (κ_i). Let us quantify this scatter by a multiplicative error function for each period (y_i and Y_i):

$$c_{ij} = \rho_j/(\kappa_i y_i) \qquad (3)$$

$$C_{il} = P_l/(\kappa_i Y_i). \qquad (4)$$

These two equations are just statements that the count reported by each observer is just the real count divided by the real K-coefficient. Both y_i and Y_i are drawn from the same distribution with the properties that the values are always positive, their average is unity, and there is some width to the distribution. This can be quantified by the following constraints on the distribution functions $D(y_i)$ and $D(Y_i)$:

$$D(y_i) = D(Y_i), \qquad (5)$$

$$y_i > 0, \langle y_i \rangle_i = 1, \langle (y_i-1)^2 \rangle_i > 0 \qquad (6)$$

$$Y_i > 0, \langle Y_i \rangle_i = 1, \langle (Y_i-1)^2 \rangle_i > 0. \qquad (7)$$

Let us also assume that for the first period the reported AAVSO number is exactly perfect, so that $r_j = \rho_j$. Alternatively, this assumption can also be stated as taking the previous epoch to define the scale for the subsequent numbers. Then, we can evaluate the K-coefficient to be used in the current period as

$$K_i = 10^{\langle \log(\rho_j/c_{ij}) \rangle_j} = 10^{\langle \log(\kappa_i y_i) \rangle_j} = 10^{\log(\kappa_i y_i)} = \kappa_i y_i, \qquad (8)$$

with the use of equations (2) and (3).

The variation in the K-coefficients can be evaluated by various means. I have analyzed the reported yearly values for more than 70 observers, many of whom contribute to the AAVSO number. A typical year-to-year variation in the K-coefficient is 10%.

Even though the published number for the previous epoch is perfect (by assumption), the current epoch will have slight deviations due to inexact measurements of the K-coefficients. Let us be quantitative in relating the published AAVSO number (R_1) and the true number (P_1) for the current period. From equations (1), (4), and (8):

$$R_1 = \langle \kappa_i y_i P_1 / (\kappa_i Y_i) \rangle_i = P_1 \langle y_i / Y_i \rangle_i. \qquad (9)$$

Now for any distribution that satisfies equations (6) and (7), I can prove (see Appendix) that

$$\langle y_i / Y_i \rangle_i > 1. \qquad (10)$$

Combining equations (9) and (10), we get

$$R_1 > P_1. \qquad (11)$$

What is the implication of equation (11)? It means that the current epoch will have an error such that the reported value is always larger than the correct value.

In general, there will be some factor (f) by which the current value will be too large:

$$R_1 = P_1 f. \qquad (12)$$

We will see later that $f \sim 1.003$. This formula will apply to every pair of time periods with a re-evaluation of the K-coefficients. So if the zeroth period has a correct number (perhaps by definition), then the Nth period will have a published value that is larger as

$$R_1(N^{th} \text{ period}) = P_1 f^N. \qquad (13)$$

Equation (13) shows that the published AAVSO number has an incremental error that increases with time such that it exponentially grows larger than the true number. This is inflation.

So how much in error is the current AAVSO number? Shapley (1949) started the current system with the Zurich numbers as a base in 1944–45 to calculate the first set of K values in 1946. Subsequent recalculations of K are made annually for any year where at least half the months have a final R >100 (Shapley 1949; Shapley 1995; Taylor 1985; Taylor 1991; Mattei 1996). However, during the years when C. Hossfield was Chairman of the AAVSO Solar Division (1964–1981), the K-coefficients were recalculated only once (Hossfield 1996). With this information, I can look over the monthly AAVSO numbers and determine the number of years with at least half the months with R > 100. Thus, N = 18, as close as can be estimated. Then the inflation to date is f^N, or 6% for $f \sim 1.003$. From 1945 to present, this total inflation is within the typical monthly scatter, and so is not yet prominent. In subsequent decades, the AAVSO number will have increasing errors, since these rise exponentially with time.

What is the root cause of this inflation? There are two needed conditions: First, the AAVSO procedure uses observations from *previous* periods to establish K for the *current* period. Second, the AAVSO uses a *linear* relation for equation (1). With both of these conditions, the small random changes in K will cause the measured K-coefficients as well as the published number to creep up with a steady inflation.

3. Realistic inflation

The above derivation of inflation has ignored the possibility of daily observational errors by each observer. Let us quantify these daily errors by multiplicative factors X_{il} and x_{ij}, which satisfy conditions similar to equations (5)–(7). Now equations (3) and (4) can be expressed as

$$c_{ij} = (x_{ij}\rho_j)/(\kappa_i y_i) \tag{14}$$

$$C_{il} = (X_{il}P_l)/(\kappa_i Y_i). \tag{15}$$

Similarly to equation (8), with $r_j = \rho_j$, we get

$$K_i = 10^{\langle \log(\rho_j/c_{ij}) \rangle_j} = \kappa_i y_i (\Pi_j x_{ij}^{-1/J}), \tag{16}$$

where there are J days of observations in the previous period. The Π_j symbol means a product over all values of j. Now from equations (1), (15), and (16):

$$R_l = \langle \kappa_i y_i (\Pi_j x_{ij}^{-1/J})(X_{il}P_l)/(\kappa_i Y_i) \rangle_i = P_l \langle y_i/Y_i \rangle_i \langle X_{il}/(\Pi_j x_{ij}^{1/J}) \rangle_i. \tag{17}$$

From a later derivation (see Appendix), we have

$$\langle X_{il}/(\Pi_j x_{ij}^{1/J}) \rangle_i > 1, \tag{18}$$

so we get equation (11) (and equations (12) and (13)) again. Thus, individual daily random errors do not change the conclusion that the AAVSO number suffers inflation.

The above derivations have assumed that all observers are equally weighted, whereas the procedure is to form a weight from the average logarithmic deviations from the scaled counts. The math then becomes tediously long, so I will not present the corresponding proof of inflation. Instead, I can report on a complete Monte Carlo simulation of the AAVSO procedure, with individual observer weights, daily observational errors, and random daily clouds. My models show inflation in all cases. In fact, the inflation factors are always identical to the case with equal weighting and no clouds. For typical K variations of 10%, observer uncertainties of 10%, and 40% cloudiness, for 50 observers, I find an average inflation of $f \approx 1.003$.

4. How to repair the AAVSO number

Inflation can be conquered in any of several ways. First, the AAVSO might abandon the current K(10G+F) formulation which leads to so many troubles and only adds noise. In particular, empirical (Hoyt *et al.* 1994) and theoretical (Schaefer 1993) arguments show that the correct formulation is simply to do a group number. However, this elegant solution would have problems with continuity.

Second, the old published AAVSO numbers must be repaired. This could be done with appropriate studies on the average variation of K-coefficients and on average observational errors. Then, detailed studies can evaluate the inflation factor. If the dates of every K calculation can be determined, then the published AAVSO numbers can be statistically deflated without examining the individual raw data.

Third, the incoming new data can be analyzed with some method that does not suffer inflation. The most obvious such method is to modify equation (1) to be a logarithmic average instead of a linear average. Thus:

$$\log(R_l) = \langle \log(K_i) + \log(C_{il}) \rangle_i . \tag{19}$$

Intermediate results then reproduce equation (8) and

$$R_l = P_l \, 10^{\langle \log(y_i/Y_i) \rangle_i} . \tag{20}$$

With

$$\langle \log(y_i/Y_i) \rangle_i = 0, \tag{21}$$

(see Appendix), we see that there is no inflation for this analysis procedure (that is, f = 1). If the old raw data are available, it might be possible to repair *post facto* the inflation in the published numbers.

5. Summary

The AAVSO number suffers from a subtle mathematical mistake which automatically increases the published number by a factor of ~1.003 every time the K-coefficients are recalculated. The root of this problem is that the observations are averaged linearly and that K-coefficients are calculated with old data. Note that the existence of inflation in the AAVSO number is a simple mathematical consequence of the AAVSO procedure. This means that there is no uncertainty concerning the existence of inflation (just as Euclid's theorems can be proven with no uncertainty).

The effect is to cause an inflation in the AAVSO numbers that causes significant rises on a time scale of decades. For short time-scale analyses, this inflation can safely be neglected. However, awareness of such systematic errors is vital for any interpretation of long-term secular changes in the Sun. For example, a spurious increase in solar activity could be attributed as the cause of the observed global warming, with profound implications for public policy and public health.

The solutions to this problem are simple and easy: first, deflate the old published data, and second, change from a linear to a logarithmic average of scaled counts. Also fundamental is to let researchers know that the old published AAVSO numbers do have inflation.

References

Hossfield, C. H. 1996, private communication.

Hoyt, D. V., Schatten, K. H., and Nesmes-Ribes, E. 1994, *Geophys. Res. Letters*, **21**, 2067.

Mattei, J. A. 1996, private communication.

Schaefer, B. E. 1993, *Astrophys. J.*, **411**, 909.

Shapley, A. H. 1949, *Pub. Astron. Soc. Pacific*, **61**, 13.

Shapley, A. H. 1995, private communication.

Taylor, P. O. 1985, *J. Amer. Assoc. Var. Star Obs.*, **14**, 28.

Taylor, P. O. 1991, *Observing the Sun*, Cambridge Univ. Press.

Appendix: Mathematical details

This appendix will prove equations (10), (18), and (21), given conditions in equations (5)–(7).

First, let me demonstrate equation (10) for reasonable distributions. I approached this by assuming various distributions for $D(y_i) = D(Y_i)$ and calculating $\langle y_i/Y_i \rangle_i$. Any square distribution, any pairs of delta functions, and any parabolic distribution can be solved analytically to reveal that the mean value is always greater than unity, while a Gaussian distribution can be evaluated by a Monte Carlo simulation with the same result. With four widely different distributions all yielding equation (10) for all second moments, it seems reasonable to adopt equation (10) in practice.

Nevertheless, let me give a general proof for any distribution that satisfies equations (6) and (7). Let us expand the mean as integrals over the distributions:

$$\langle y_i/Y_i \rangle_i = \int dy_i \int dY_i \, D(y_i)\, D(Y_i)\, (y_i/Y_i). \qquad (22)$$

This can be separated as

$$\langle y_i/Y_i \rangle_i = [\int dy_i \, D(y_i)\, y_i]\, [\int dY_i \, D(Y_i) / Y_i]. \qquad (23)$$

The term in the first square bracket is simply $\langle y_i \rangle_i$, which is 1 from equation (6). So

$$\langle y_i/Y_i \rangle_i = \int dY_i \, D(Y_i) / Y_i = \langle 1/Y_i \rangle_i . \qquad (24)$$

Now consider the expression $(Y_i-1)^2/Y_i$. From equation (7), we know that both the numerator and the denominator are greater than zero, so their ratio must always be greater than zero, so the average ratio must be greater than zero:

$$\langle (Y_i-1)^2/Y_i \rangle_i > 0. \qquad (25)$$

The expression can be expanded as

$$\langle (Y_i-1)^2/Y_i \rangle_i = \langle Y_i \rangle_i + \langle (-2) \rangle_i + \langle 1/Y_i \rangle_i = 1 + (-2) + \langle 1/Y_i \rangle_i \qquad (26)$$

with the $\langle Y_i \rangle_i$ term evaluated from equation (7). Now combining equations (24), (25), and (26), we find

$$\langle y_i/Y_i \rangle_i > 1, \qquad (27)$$

which is the same as equation (10).

Let me offer a second proof of equation (10). From equation (22), we can rename variables being integrated over to find

$$\langle y_i/Y_i \rangle_i = \langle Y_i/y_i \rangle_i . \qquad (28)$$

Simple manipulation of this equation gives

$$\langle y_i/Y_i \rangle_i = \langle (y_i/Y_i + Y_i/y_i)/2 \rangle_i . \qquad (29)$$

Now, consider the quantity $Q \equiv (A/B + B/A)/2$ for any positive values of A and B. The minimum occurs when $\partial Q/\partial A = 0$ and $\partial Q/\partial B = 0$ (with the second partial derivatives positive). The smallest possible value of Q is unity when $A = 1$ and $B = 1$. So $Q > 1$ as long as A and B are not both exactly equal to unity. Now let us identify y_i with A and Y_i with B. Then $(y_i/Y_i + Y_i/y_i)/2$ will always be greater than unity as long as either y_i or Y_i do not both equal unity. If we average this quantity over the distributions $D(Y_i)$ and $D(y_i)$, then we will be averaging 1 with values greater than 1 (since the positive second moments [equations (6) and (7)] ensures that at least some values will have

$Y_i \neq 1$ or $y_i \neq 1$), so the average must be greater than unity. Thus,

$$\langle y_i/Y_i \rangle_i = \langle (y_i/Y_i + Y_i/y_i)/2 \rangle_i > 1, \tag{30}$$

which is equation (10) again.

The proof for equation (18) is similar to this second proof of equation (10). The random variable X_{il} can be expanded as $\Pi_j X_{il}^{1/J}$. Now the average in equation (18) can be expressed as

$$\langle X_{il}/\Pi_j x_{ij}^{1/J} \rangle_i = \langle \Pi_j (X_{il}/x_{ij})^{1/J} \rangle_i = \Pi_j \langle (X_{il}/x_{ij})^{1/J} \rangle_i. \tag{31}$$

The averages in the last term are all the same, so

$$\langle X_{il}/(\Pi_j x_{ij}^{1/J}) \rangle_i = (\langle (X_{il}/x_{ij})^{1/J} \rangle_i)^J. \tag{32}$$

As in equation (28), we can change integration variable names without effect since the distributions are the same, so that

$$\langle (X_{il}/x_{ij})^{1/J} \rangle_i = \langle (x_{ij}/X_{il})^{1/J} \rangle_i. \tag{33}$$

Simple manipulation then gives

$$\langle (X_{il}/x_{ij})^{1/J} \rangle_i = \langle [(x_{ij}/X_{il})^{1/J} + (X_{il}/x_{ij})^{1/J}]/2 \rangle_i. \tag{34}$$

Now let us consider the quantity $Q \equiv [(A/B)^{1/J} + (B/A)^{1/J}]/2$, for any positive A, B, and J. The minimum occurs when $\partial Q/\partial A = 0$ and $\partial Q/\partial B = 0$ (with the second partial derivatives positive). This minimum occurs when $A = B$, in which case the minimum Q value is 1. So $Q > 1$ as long as $A \neq B$. Now let us identify x_{ij} with A and X_{il} with B. Then the average in the right side of equation (34) is of quantities that are equal to unity (if $x_{ij} = X_{il}$) and greater than unity (if $x_{ij} \neq X_{il}$). The non-zero second moments of the distribution force the case with $x_{ij} \neq X_{il}$ to occur in the average, so the average must be greater than unity:

$$\langle [(x_{ij}/X_{il})^{1/J} + (X_{il}/x_{ij})^{1/J}]/2 \rangle_i > 1. \tag{35}$$

With equations (32), (34), and (35), we find

$$\langle X_{il}/(\Pi_j x_{ij}^{1/J}) \rangle_i > 1, \tag{36}$$

which is just equation (18).

Now to prove equation (21). First let us explicitly write integrals for the means

$$\langle \log(y_i) \rangle_i = \int dy_i \log(y_i) D(y_i), \tag{37}$$

$$\langle \log(Y_i) \rangle_i = \int dY_i \log(Y_i) D(Y_i) = \int dy_i \log(y_i) D(y_i) = \langle \log(y_i) \rangle_i. \tag{38}$$

In the middle part of the last equation, we renamed the variable being integrated over and used equation (5). Now,

$$\langle \log(y_i/Y_i) \rangle_i = \langle \log(y_i) \rangle_i - \langle \log(Y_i) \rangle_i = 0. \tag{39}$$

This proves equation (21).