Solar Bulletin

THE AMERICAN ASSOCIATION OF VARIABLE STAR OBSERVERS
SOLAR SECTION

Rodney Howe, Editor, Chair
c/o AAVSO, 49 Bay State Rd
Cambridge, MA 02138

Web: http://www.aavso.org/solar-bulletin
Email: solar@aavso.org
ISSN 0271-8480

Volume 74 Number 2 February 2018

The Solar Bulletin of the AAVSO is a summary of each month's solar activity recorded by visual solar observers' counts of group and sunspots and the VLF radio recordings of SID Events in the ionosphere. Section 1 gives contributions by our members. The sudden ionospheric disturbance report is in Section 2. The relative sunspot numbers are in Section 3. Section 4 has endnotes.

1 Statistical Models for Counts Data

The common practice of manipulating counts data to approximate a Gaussian distribution is prone to several shortcomings. For example, counts data range from zero to infinity, and a Gaussian probability distribution (PDF) function which extends between $\pm\infty$ becomes skewed. The skew gives the appearance outliers of large count values. The mean of the counts shifts as a function of the variance of the counts; i.e., as the variance increases, so does the mean. This is contrary to the Gaussian distribution which is characterized by a mean that is independent of the magnitude of the variance. Applying a log transformation to counts data will compress these data, but it usually does not contract the skew enough to allow for an approximation of a Gaussian distribution. If the skew remains post transformation, so-called outliers are still present. Removing them simply changes the shape of the compression, and "new" outliers appear. This compression phenomenon is seen with Wolf numbers, which are a combination of counts of sunspots and of sunspot groups.

Counts models used by statisticians are part of a broad class of models called generalized linear models (McCullagh and Nelder [1989]). Generalized linear models (GLMs) are specified [Agresti, 1998] by three components: a random component that identifies the counts variable probability distribution which, for Wolf numbers, usually is the Poisson, the quasi-Poisson, or the negative binomial PDF; a systematic component consisting of variables that affect counts, which includes observer designators, the date and time of the observations, the seeing conditions, and the method used to count sunspots, among others for Wolf numbers; and a function called a link that specifies the relationship between the expected value of the Wolf number and the systematic component, e.g., a natural log transformation. A simple example of a model for Wolf numbers R_a is

$$\begin{aligned} \mu &= e^{\beta_0}\, e^{\beta_1 x_1}\, e^{\beta_2 x_2} \\ &= e^{\beta_0+\beta_1 x_1+\beta_2 x_2} \\ \Rightarrow \log\mu &= \beta_1 x_1 + \beta_2 x_2. \end{aligned} \tag{1}$$

where μ is the expected value of the Wolf number as generated by the model predictor variables x_1 = observer, say, and x_2 = observing method such as projection or H-alpha filtering. The model estimation process calculates the values of β_0 (the model intercept), β_1, and β_2, estimated by likelihood methods, which connect the expected value of the Wolf number to the example's observer and corresponding observation method predictors.

It is important to note that the Wolf number (R_a) data are not transformed as $\log(R_a)$, rather the expected value of the Wolf number as given by a multiplication of exponential functions is transformed to yield the log of the mean as an additive relationship using the natural log link function as in the example.

The simple model example above has only two predictor variables, observer and observing method, that affect the expected R_a. Current Wolf number modeling includes seeing condition, magnification, objective size, telescope type, and filter type. We test these variables to understand which if any affect the variation in the expected value of the Wolf number. The GLM thereby accounts for the sources of Wolf number variability and uses only the "leftover" variance to establish a confidence interval on the estimated monthly Wolf number. The GLM gives a more accurate estimate as only the error in the count remains, the other sources of error having been identified and adjusted out of the count error.

The Shapley approach to modeling Wolf numbers (used by the AAVSO) depends upon a sufficient transformation of counts data to follow a Gaussian distribution. The transformation is required to force homogeneous variance from the lowest counts to the largest, though there are falsely identified outliers which, if a counts distribution is used, are not so designated. Biased Wolf numbers result when outliers are assumed, and are removed to obtain a stable transformation. Thus, information contained by the outliers is lost to the analysis. In addition, biased model parameters under-estimate the leftover error of the model which often assigns significance to model coefficients which would be otherwise benign.

The GLM is specifically designed to model Wolf numbers that follow a counts distribution. (The evolution of the GLM may be found in Nelder and Wedderburn [1972], Wedderburn [1972], Nelder and Pregibon [1987], Lee and Nelder [1996] and Lee and Nelder [2001].) No data need be removed because the counts distribution is skewed in the direction of larger counts, and no zero counts need be removed. The error structure used to model the Wolf number distribution accounts for the mean of the counts is not equaling the variance of the counts and hence no information is eliminated due to the thick, right-tailed behavior of the counts distribution. Dissimilar variance structure modeling in GLMs leads to correct determination of unbiased model parameter estimation and significance. Modern GLM construction produces monthly Wolf number estimates that are more efficient and consistent than the Shapley method with assumed Gaussian-distributed data. See Riggs and Lalonde [2017] for further information on counts models, and http://www.spesi.org/?page_id=65 for more information on the statistical modeling of Wolf numbers.

Unfortunately, a most accurate estimate of the monthly Wolf number is not currently possible due to missing data for the predictor variables. The best estimate can be made after each of you observers supply these data if they are missing from your file header.

So, please spend a few minutes to update your SunEntry Header data, both current and historical, with complete and accurate information. This will help with creating a more robust GLM model for the American Relative index.

2 Sudden Ionospheric Disturbance (SID) Report

Here we show how a 24 bit external sound card can be used to record VLF SID data without a receiver or any electric amplification: (https://www.asus.com/us/Sound-Cards/Xonar_U5/) And a Pi Zero computer running Linux (Jessie) operating system.

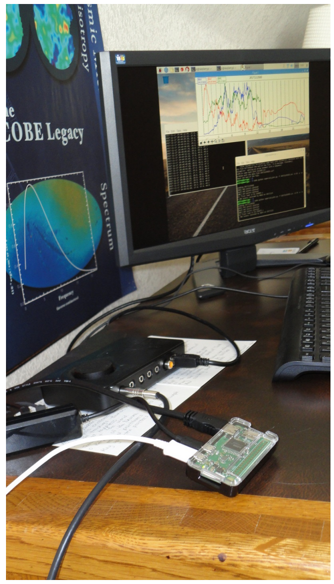

Figure 1: The Xonar external sound card (black box) and Pi Zero computer (foreground)

2.1 SID Records

February 2018 (Figure 2) There were 15 SID events recorded on the 11th of February here in Fort Collins, Colorado.

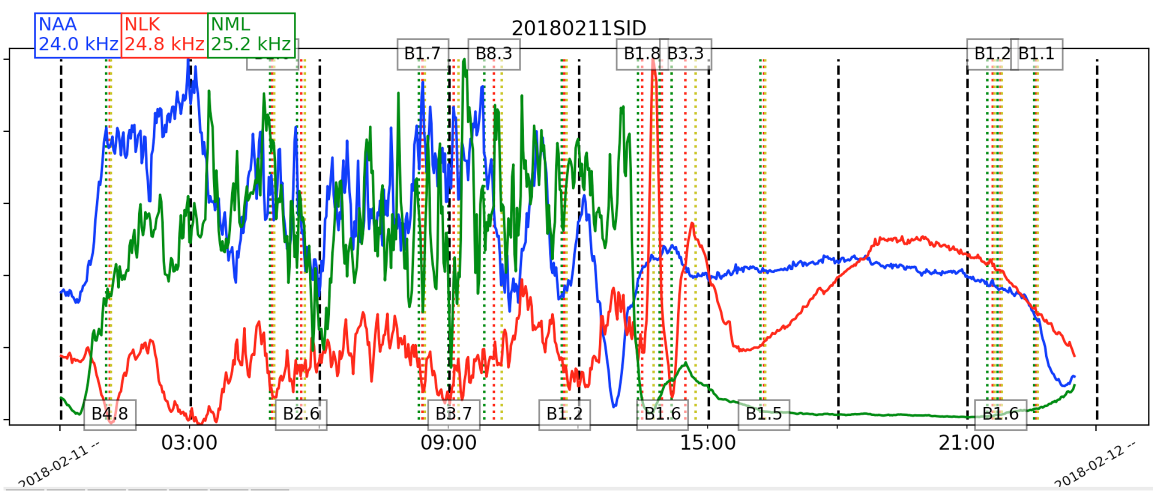

Figure 2: VLF recording using the sidmon.py software from Nathan Towne.

2.2 SID Observers

In February 2018 we have 15 AAVSO SID observers who submitted VLF data as listed in Table 1. Observers monitor from one to three stations to provide SID data.

Table 1: 201802 VLF Observers

Observer	Code	Stations
A McWilliams	A94	NML
R Battaiola	A96	ICV
J Wallace	A97	NAA
L Loudet	A118	GBZ
J Godet	A119	GBZ GQD ICV
B Terrill	A120	NWC
F Adamson	A122	NWC
S Oatney	A125	NML
J Karlovsky	A131	DHO NSY
R Green	A134	NWC
R Mrllak	A136	GQD NSY
S Aguirre	A138	NPM
R Rogge	A143	DHO GQD ICV
K Menzies	A146	NAA
L Ferreira	A149	NWC

Figure 3 depicts the importance rating of the solar events. The durations in minutes are -1: LT 19, 1: 19-25, 1+: 26-32, 2: 33-45, 2+: 46-85, 3: 86-125, and 3+: GT 125.

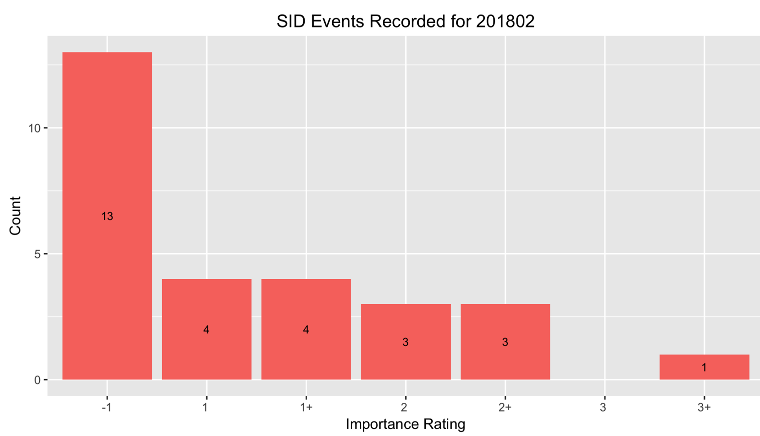

Figure 3: Solar Events Y-axis, Importance Rating X-axis.

2.3 Solar Flare Summary from GOES-15 Data

In February 2018, There were 94 solar flares measured by GOES-15. Six C class, 86 B class flares and 2 A class flares. A lot more flaring this month compared to last month. There were 12 days this month with no GOES-15 reports of flares. (see Figure 4).

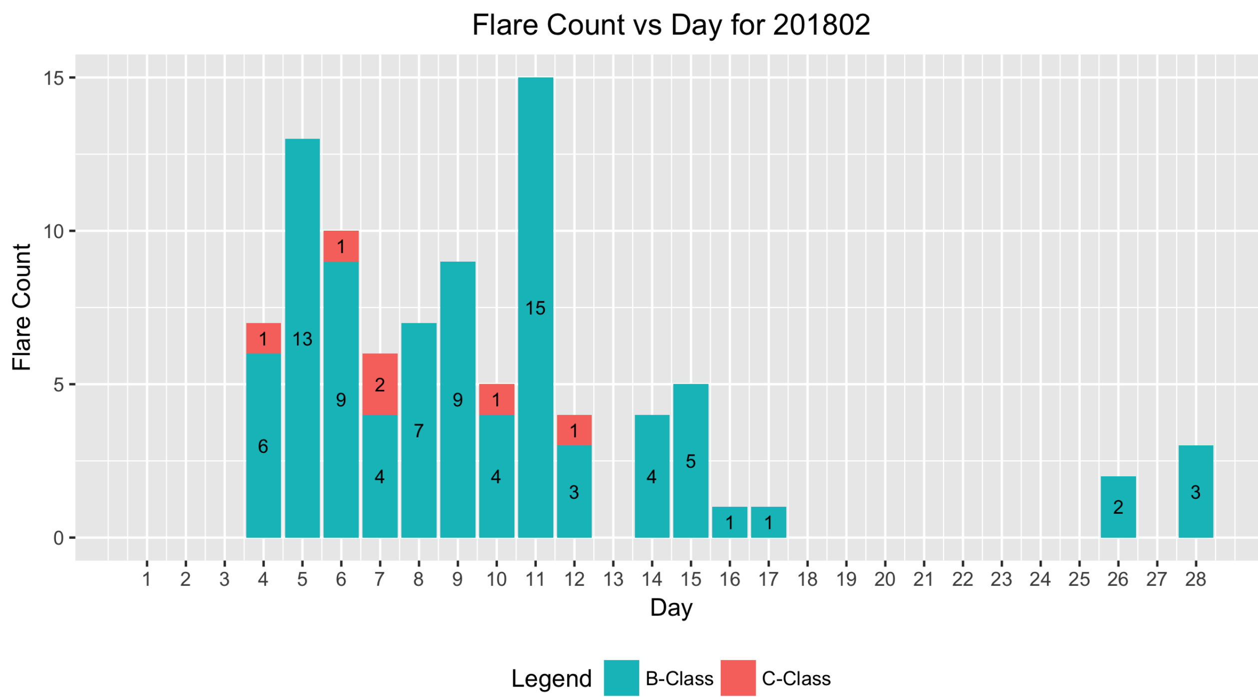

Figure 4: GOES - 15 XRA flares

3 Relative Sunspot Numbers (Ra)

Reporting monthly sunspot numbers consists of submitting an individual observer's daily counts for a specific month to the AAVSO Solar Section. These data are maintained in a SQL database. The monthly data then are extracted for analysis. This section is the portion of the analysis concerned with both the raw and daily average counts for a particular month. Scrubbing and filtering the data assure error-free data are used to determine the monthly sunspot numbers.

3.1 Raw Sunspot Counts

The raw daily sunspot counts consist of submitted counts from all observers who provided data in February 2018. These counts are reported by the day of the month, and are either from data not scrubbed or corrected data.

The reported raw daily average counts have been checked for errors and inconsistencies, and no known errors are present. All observers whose submissions qualify through this month's scrubbing process are represented in Figure 6.

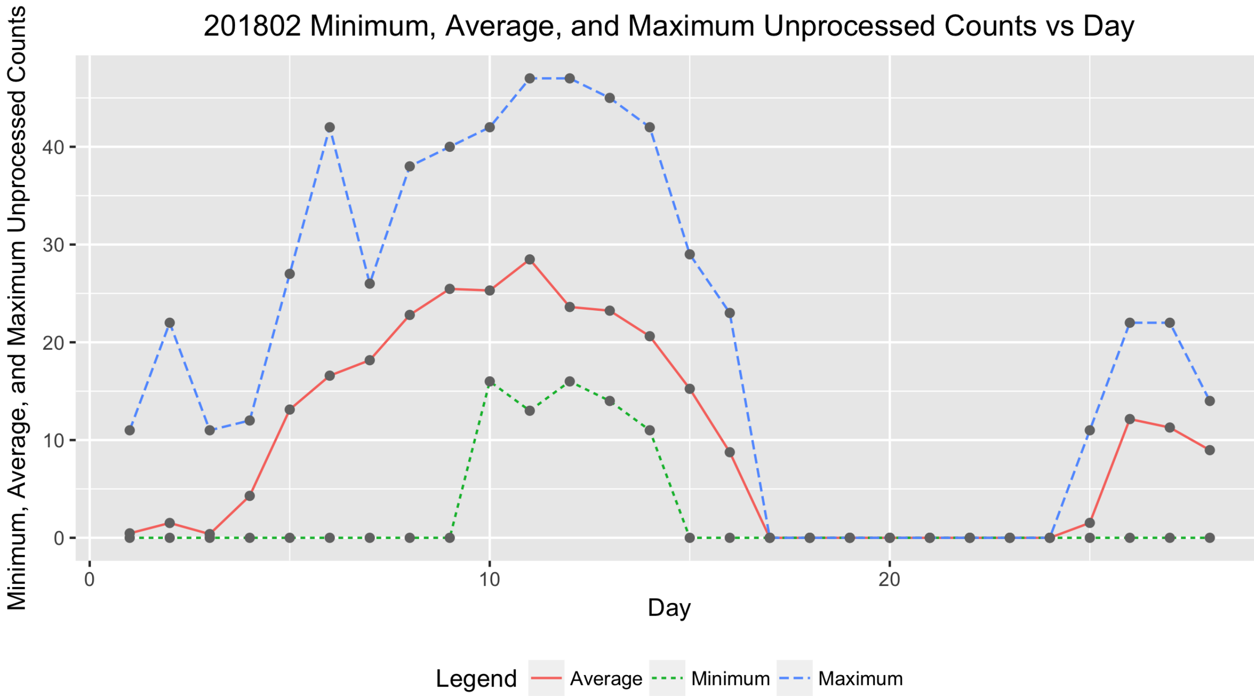

Figure 5: Raw average, minimum and maximum counts by day of the month by observer.

3.2 American Relative Sunspot Numbers

The relative sunspot numbers, R_a contain the sunspot numbers after the submitted data are scrubbed and modeled by Shapley's method with k-factors (http://iopscience.iop.org/article/10.1086/126109/pdf). The Shapley method is a statistical model that agglomerates variation due to random effects such as observer and fixed effects such as seeing condition. See Table 2.

Table 2: 201802 American Relative Sunspot Numbers (Ra)

Day	NumObs	Raw	Ra
1	24	0	0
2	29	0	0
3	29	0	0
4	31	5	4
5	34	10	8
6	29	14	10
7	30	17	11
8	35	21	14
9	28	24	16
10	20	26	20
11	25	30	21
12	31	24	16
13	34	26	16
14	29	21	13
15	29	15	10

Continued

Table 2: 201802 American Relative Sunspot Numbers (Ra)

Day	NumObs	Raw	Ra
16	29	10	5
17	33	0	0
18	33	0	0
19	20	0	0
20	28	0	0
21	28	0	0
22	29	0	0
23	31	0	0
24	38	0	0
25	36	1	1
26	34	12	8
27	39	14	9
28	31	12	8
Averges	30.2	10.1	6.8

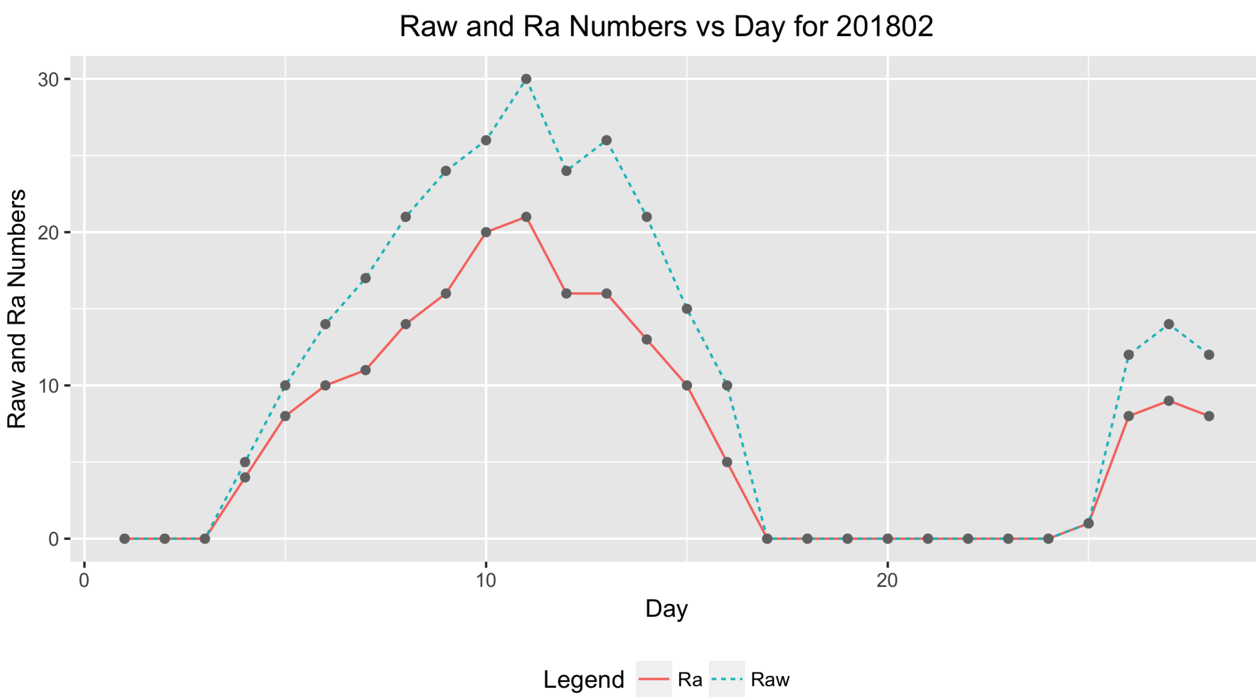

Figure 6: Raw Wolf and Ra numbers by day of the month by observer.

3.3 Sunspot Observers

Table 3 lists the observer code (obs), the number of observations submitted for February 2018, and the observer's name. The final rows of the table give the total number of observers who submitted sunspot counts and the total number of observations submitted. The total number of observers is 63 and the total number of observations is 846.

Table 3: 201802 Number of observations by observer

Obs	NumObs	Name
AAX	19	Alexandre Amorim
AJV	12	J. Alonso
ARAG	28	Gema Araujo
ASA	14	Salvador Aguirre
BARH	10	Howard Barnes
BERJ	18	Jose Alberto Berdejo
BMF	14	Michael Boschat
BRAD	28	David Branchett
BRAF	12	Raffaello Braga
BROB	24	Robert Brown
BSAB	23	Santanu Basu
CHAG	20	German Morales Chavez
CIOA	16	Ioannis Chouinavas
CKB	9	Brian Cudnik
CNT	14	Dean Chantiles
CVJ	6	Jose Carvajal
DEMF	2	Frank Dempsey
DJOB	11	Jorge del Rosario
DMIB	14	Michel Deconinck
DROB	8	Bob Dudley
DUBF	24	Franky Dubois
ERB	11	Bob Eramia
FERJ	14	Javier Ruiz Fernandez
FLET	12	Tom Fleming
FLF	7	Fredirico Luiz Funari
FTAA	8	Tadeusz Figiel
FUJK	21	K. Fujimori
HAYK	7	Kim Hay
HMQ	3	Mark Harris
HOWR	21	Rodney Howe
JDAC	6	David Jackson
JENS	1	Simon Jenner
JGE	6	Gerardo Jimenez Lopez
JPG	5	Penko Jordanov
KAPJ	11	John Kaplan
KNJS	28	James & Shirley Knight
KROL	16	Larry Krozel
LEVM	15	Monty Leventhal
LKR	3	Kristine Larsen
LRRA	11	Robert Little
MCE	25	Etsuiku Mochizuki
MILJ	9	Jay Miller

Continued on next page

Table 3: 201802 Number of observations by observer

Obs	NumObs	Name
MJAF	27	Juan Antonio Moreno Quesada
MJHA	25	John McCammon
MUDG	8	George Mudry
MWU	18	Walter Maluf
ONJ	5	John O'Neill
RLM	12	Mat Raymonde
SDOH	28	Solar Dynamics Obs - HMI
SIMC	5	Clyde Simpson
SMNA	3	Michael Stephanou
SNE	1	Neil Simmons
SONA	14	Andries Son
SPIA	6	Piotr Skorupski
STAB	25	Brian Gordon-States
SUZM	23	Miyoshi Suzuki
TESD	16	David Teske
TPJB	3	Patrick Thibault
URBP	15	Piotr Urbanski
VARG	26	A. Gonzalo Vargas
VIDD	8	Daniel Vidican
WGI	1	Guido Wollenhaupt
WILW	11	William M. Wilson
Totals	846	63

3.4 Generalized Linear Model of Sunspot Numbers

Dr. Jamie Riggs, Solar System Science Section Head, International Astrostatistics Association, maintains a relative sunspot number (R_a) model containing the sunspot numbers after the submitted data are scrubbed and modeled by a Generalized Linear Mixed Model (GLMM), which is a different model method from the Shapley method of calculating R_a in Section 3 above. The GLMM is a statistical model that accounts for variation due to random effects and fixed effects. For the GLMM R_a model random effects include the AAVSO observer as these observers are a selection from all possible observers, and the fixed effects include seeing conditions at one of four possible levels. More details on GLMM are available in a paper (GLMM05) on http://www.spesi.org/?page_id=65 of the sunspot counts research page. The paper title is *A Generalized Linear Mixed Model for Enumerated Sunspots*.

Figure 7 shows the monthly GLMM R_a numbers for the 24th solar cycle to date. The solid cyan curve that connects the red X's is the GLMM model R_a estimates of excellent seeing conditions, which in part explains why these R_a estimates often are higher than the Shapley R_a values. The dotted black curves on either side of the cyan curve depict a 99% confidence band about the GLMM estimates. The confidence band uses the large sample approximation based on the Gaussian distribution. The green dotted curve connecting the green triangles is the Shapley method R_a numbers. The dashed blue curve connecting the blue O's is the SILSO values for the monthly sunspot numbers.

The tan box plots for each month are the actual observations submitted by the AAVSO observers. The heavy solid lines approximately midway in the boxes represent the count medians. The box plot represents the InterQuartile Range (IQR), which depicts from the 25^{th} through the 75^{th} quartiles. The lower and upper whiskers extend 1.5 times the IQR below the 25^{th} quartile, and 1.5 times the IQR above the 75^{th} quartile. The black dots below and above the whiskers traditionally are considered outliers, but with GLMM modeling, they are observations that are accounted for by the GLMM model.

4 Endnotes

Reporting Addresses

- Sunspot Reports: Kim Hay solar@aavso.org
- SID Solar Flare Reports: Rodney Howe ahowe@frii.com

References

A. Agresti. *Categorical Data Analysis*. John Wiley & Sons, Inc, New York, 3rd edition, 1998.

Y. Lee and J. A. Nelder. Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society B*, 58(4):619–678, 1996.

Y. Lee and J. A. Nelder. Hierarchical generalised linear models : A synthesis of generalised linear models, random-effect model and structured dispersion. *Biometrika*, 88:987–1006, 2001.

P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Chapman & Hall/CRC, Boca Raton, Florida, 1989.

J. A. Nelder and D. Pregibon. An extended quasi-likelihood function. *Biometrika*, 74(2):221–232, 1987.

J. A. Nelder and R. W. M. Wedderburn. Generalised linear models. *Journal of the Royal Statistical Society A*, 135:370–384, 1972.

Jamie D. Riggs and Trent L. Lalonde. *Handbook for Applied Modeling: Non-Gaussian and Correlated Data*. Cambridge University Press, 2017.

R. W. M. Wedderburn. Quasi-likelihood functions, generalized linear models, and the gauss-newton method. *Biometrika*, 61(3):439–447, December 1972.

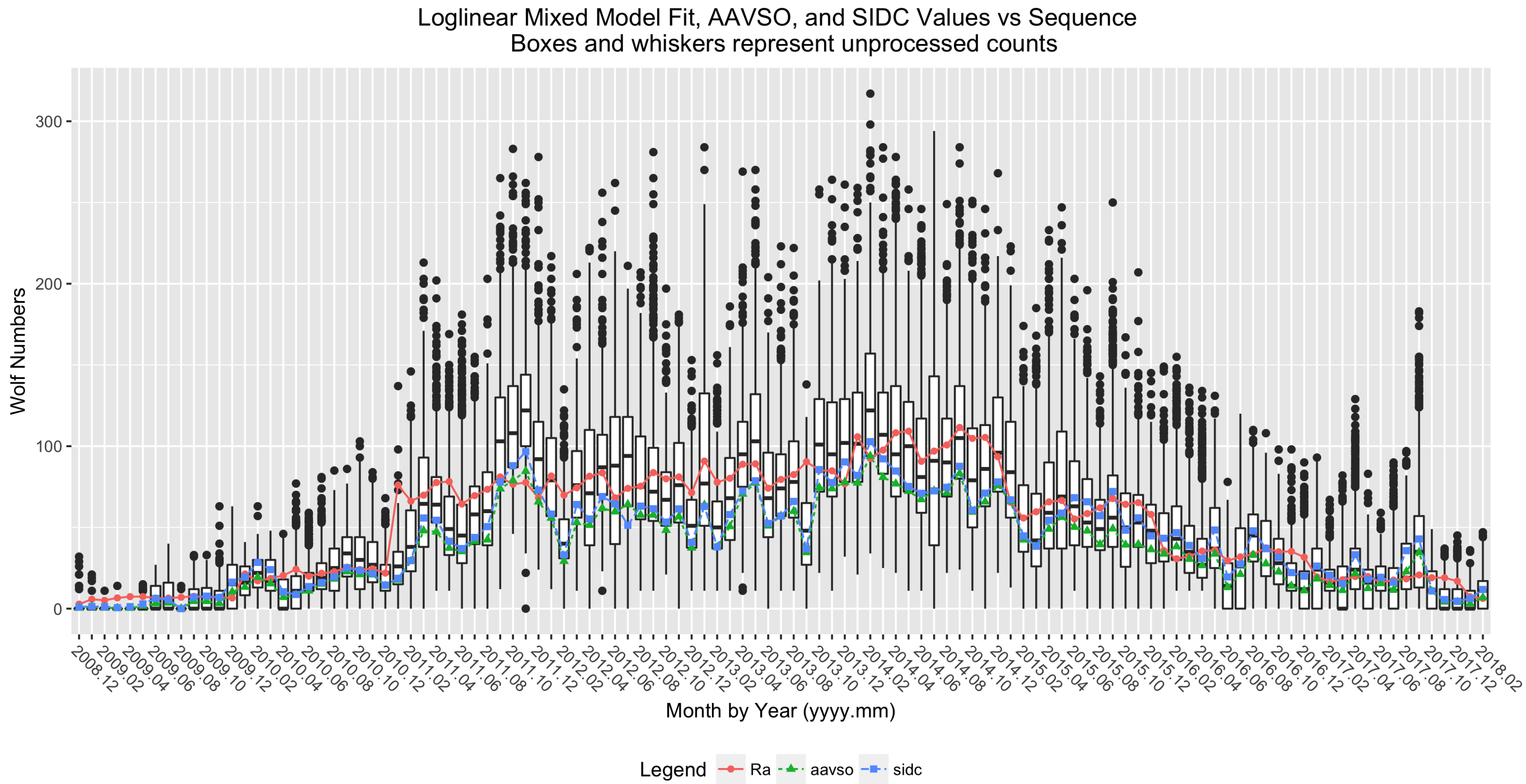

Figure 7: GLMM fitted data for R_a. AAVSO data: https://www.aavso.org/category/tags/solar-bulletin. SILSO data: WDC-SILSO, Royal Observatory of Belgium, Brussels