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 5 May 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 Wolf Number and Statistical Distributions

The Wolf number is an interesting compound representation of counts data. The Wolf number is compound as it sums group counts and spot counts as follows:

$$W = 10g + s, \tag{1}$$

where W is the Wolf number, g is the number of groups which is multiplied by 10, and s is the number of sunspots, each term on the same given day. We here discuss statistical methods for managing Wolf numbers. These methods are not necessarily concerned with preserving continuity relative to the historical Zürich numbers.

Statistical models often use a probability distribution function (pdf). Perhaps the best known pdf is the Gaussian (a.k.a normal and bell curve). We will discuss the Poisson pdf (named for Baron Siméon Denis Poisson, 21 June 1781 - 25 April 1840, a French mathematician and physicist) with a nod to the Gaussian pdf (named for Johann Carl Friedrich Gauss, 30 April 1777 - 23 February 1855, a German mathematician and physicist).

If counts from several observers follow a Poisson pdf, let's say the sunspot counts s, then the mean or average, λ_s, of these counts is equal to the variance (standard deviation squared). Conveniently, we need only one value to completely specify the counts distribution. The same mean and variance relationship can hold for the group count, g, and we denote its mean and variance by λ_g. Does this relationship hold for the Wolf number?

The Wolf number given by Eq. 1 is the sum of two Poisson pdfs. A very nice property of the Poisson pdf is that the means can be summed which gives us an expected Wolf number mean (λ_W) of

$$\lambda_W = 10\lambda_g + \lambda_s. \tag{2}$$

We may multiply λ_g by ten as it is the equivalent of summing λ_g ten times. The ability to sum the means suggests the Wolf number follows a Poisson pdf but does it?

Figure 1 is a histogram (a ranked bar graph) of the Wolf numbers (the gray bars) for from December, 2009 through May, 2018. Superimposed over the bars are a Gaussian pdf (solid black curve) and a Poisson pdf (red dotted curve). Each curve is based on the mean and variance of the Wolf number data. If either the Gaussian or the Poisson describe the pdf of the Wolf number, then either or both these curves would "ride" close to the tops of the bars. Clearly, neither curve is a good fit. Incidentally, Wolf numbers usually are "shoe-horned" into an approximate Gaussian pdf which statisticians generally consider to be poor form.

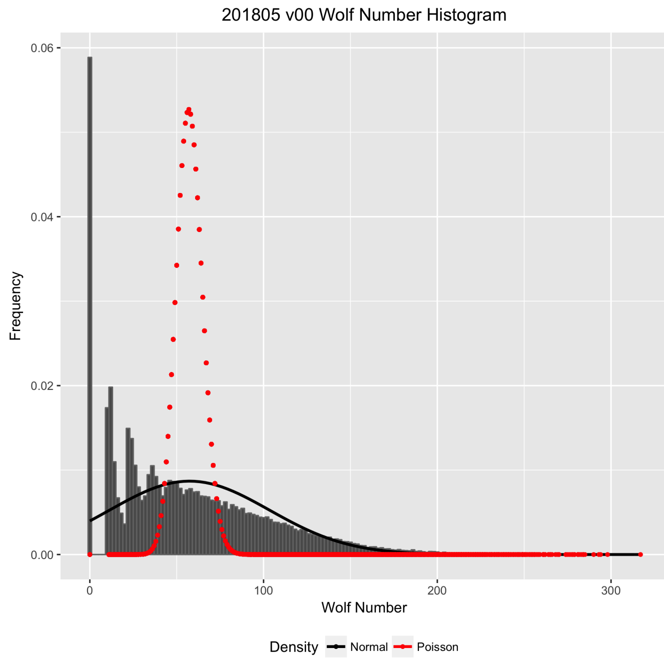

Figure 1: Wolf number histogram for data from December, 2009 through May, 2018. The bars are binned Wolf numbers, the solid black curve is a fitted normal (Gaussian) pdf, and the red dotted curve is a fitted Poisson pdf.

Why do the Wolf numbers fail to follow a Poisson pdf? From Fig. 1 we see three characteristics that contribute to the failure. The first is the inflated number of zeros (tallest bar). We see from the red dotted Poisson curve that there should be nearly no zeros though we know there must be many zeros. The second characteristic is the unusual presence of four or five "fingers" in the small Wolf numbers. Finally, the third characteristic is that the Wolf number variance (2107.958) is much, much greater than its mean (57.390). When the variance is larger than the mean, we have a condition known as overdispersion.

Fortunately, the three characteristics of Wolf number deviation from a Poisson pdf have remedies. We won't give details on these methods as each is a subject for a bulletin article. The large number of zeros is modeled by a zero-inflation model which divides the counts model into two parts; one for the zeros and the other for the non-zero counts.

The reason for the fingers is a current topic of investigation. One contributor is from the number of groups. A simple example is the contrast between one group of three sunspots (the Wolf number is $10 \times 1 + 3 = 13$), and the 3 sunspots forming three groups ($10 \times 3 + 3 = 33$. Clearly, these two results are quite different with the group term forming clusters (fingers) of Wolf numbers. Note the smallest non-zero Wolf number is 11 $10 \times 1 + 1 = 11$).

There are many pdfs that may be used to account for overdispersion in the Wolf number. Current investigation is the use of the statistical model known as a Hidden Markov Model that uses a mixture of multiple pdfs to describe all three of the deviation characteristics.

There are several levels of Wolf number investigations in progress that we hope will give new and exciting information on sunspot activity. As the research progresses, we will report results summaries in this bulletin.

2 Sudden Ionospheric Disturbance (SID) Report

2.1 SID Records

May 2018 (Figure 2) There were 9 GOES events recorded on the 28th of May. This plot from Nathan Towne in Magdalena, New Mexico records the C 2.7 flare as a SID around 1700 UT.

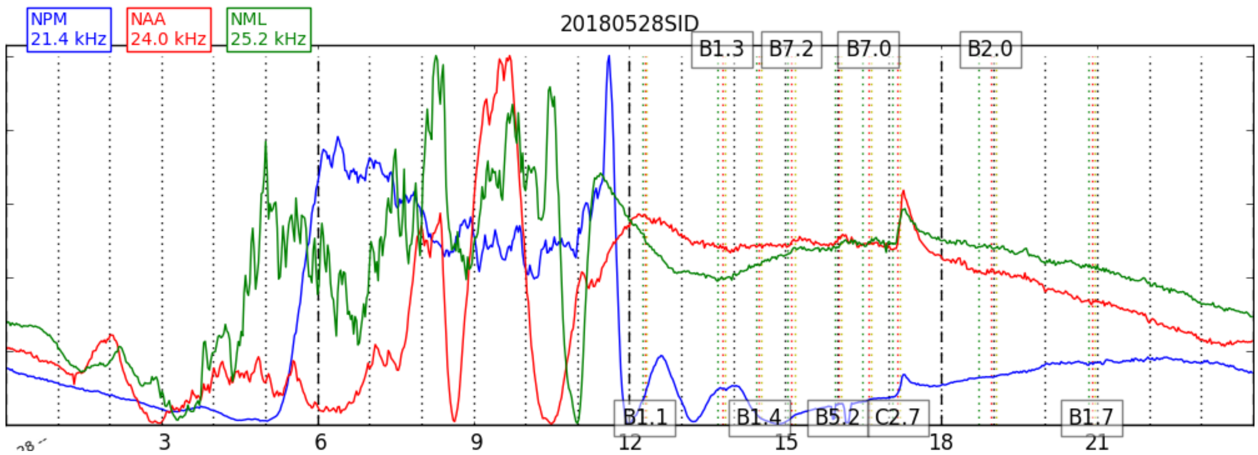

Figure 2: VLF recording using the sidmon.py software from Nathan Towne.

2.2 SID Observers

In May 2018 we have 12 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: 201805 VLF Observers

Observer	Code	Stations
A McWilliams	A94	NML
R Battaiola	A96	HWU
J Wallace	A97	NAA
L Loudet	A118	DHO
J Godet	A119	GBZ ICV
F Adamson	A122	NWC
S Oatney	A125	NML
J Karlovsky	A131	DHO ICV
S Aguirre	A138	NPM
R Rogge	A143	GQD
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.

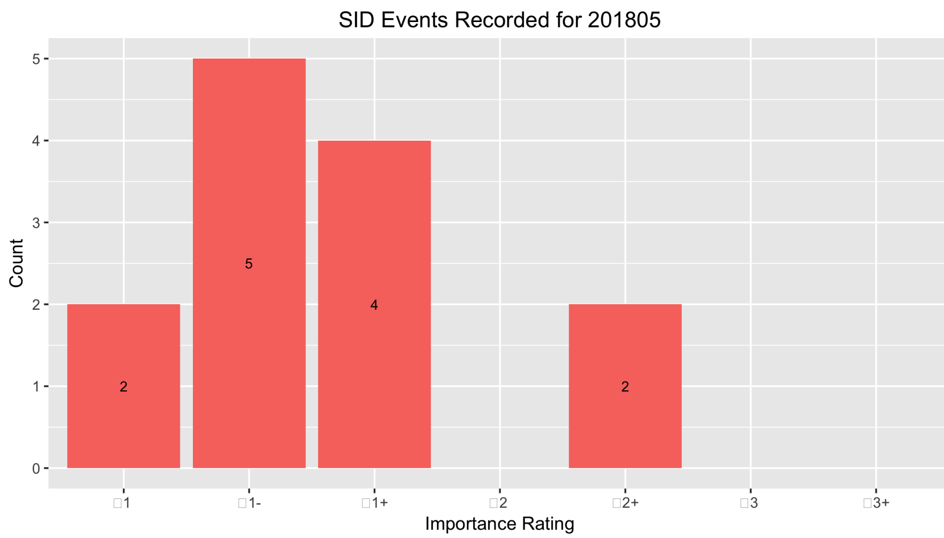

Figure 3: Solar Events Y-axis, Importance Rating X-axis.

2.3 Solar Flare Summary from GOES-15 Data

In May 2018, There were 69 solar flares measured by GOES-15: Two C class, 65 B class flares and 2 A class flares. More flaring this month compared to last month. There were 20 days this month with no GOES-15 reports of flares. (see Figure 4).

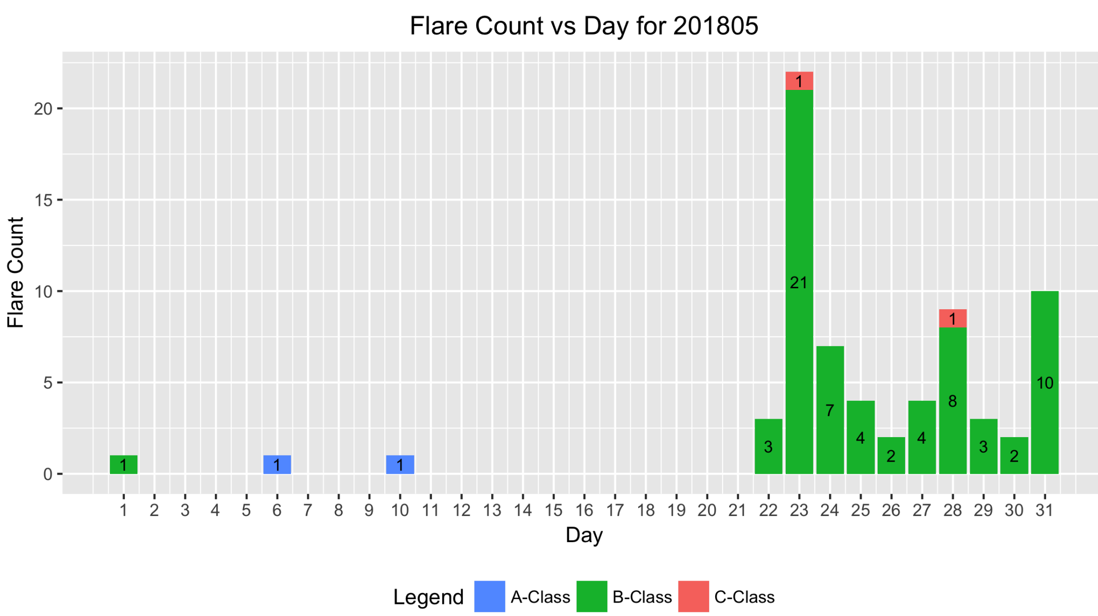

Figure 4: GOES - 15 XRA flares

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 May 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 5.

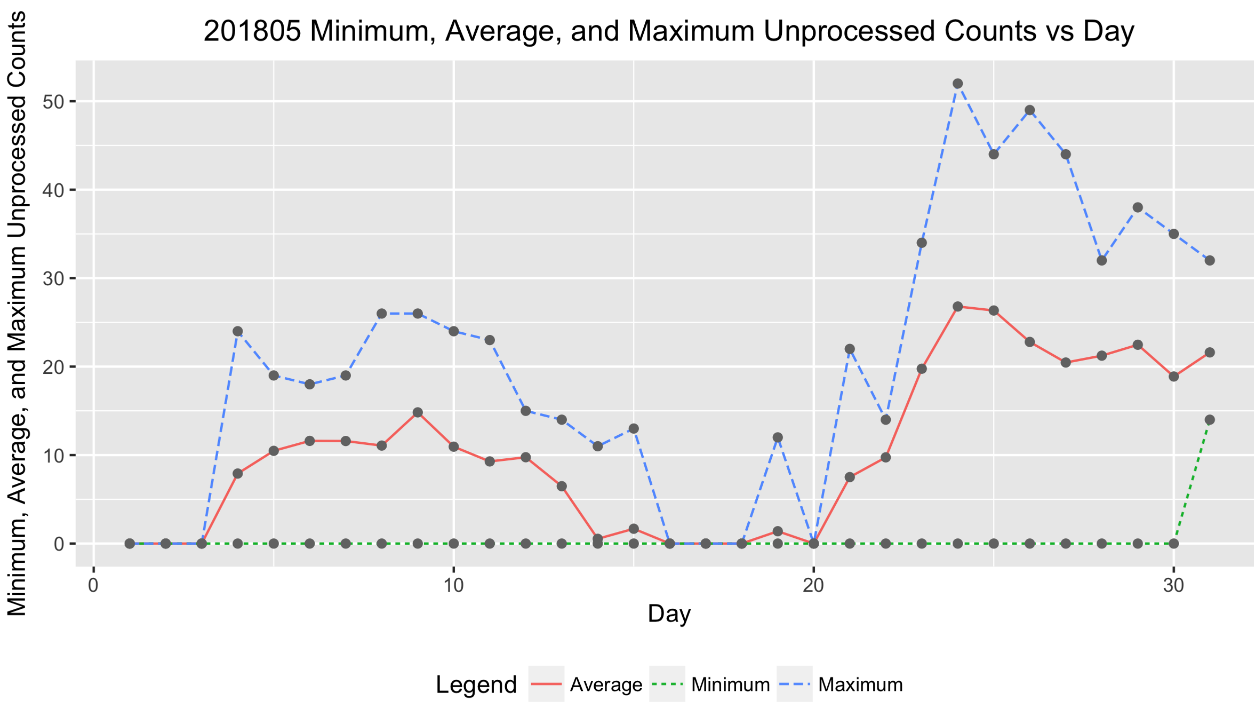

Figure 5: Raw Wolf number average, minimum and maximum by day of the month for all observers.

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: 201805 American Relative Sunspot Numbers (Ra)

Day	NumObs	Raw	Ra
1	40	0	0
2	37	0	0
3	36	0	0
4	35	10	7
5	48	14	10
6	43	13	10
7	39	14	10
8	41	12	8
9	41	18	13
10	40	13	9
11	39	11	8
12	33	12	8
13	33	10	7
14	41	1	1
15	37	2	1
16	40	0	0
17	35	0	0
18	40	0	0
19	33	3	2
20	35	0	0
21	41	6	4
22	35	11	8
23	42	19	15
24	44	29	20
25	38	30	21
26	38	25	18
27	39	23	16
28	38	23	16
29	36	23	16
30	34	21	15
31	33	24	17
Averges	38.2	11.8	8.4

3.3 Sunspot Observers

Table 3 lists the observer code (obs), the number of observations (NumObs) submitted for May 2018, and the observer's name (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 65 and the total number of observations is 1184.

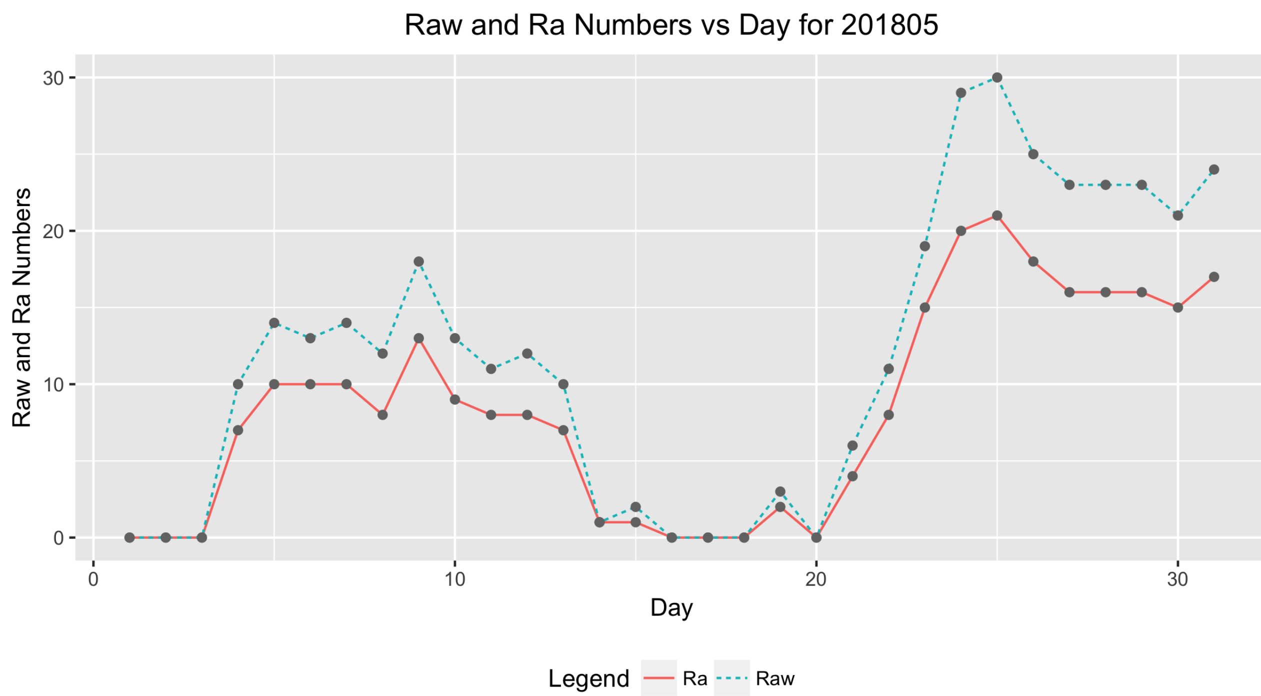

Figure 6: Raw Wolf average and Ra numbers by day of the month for all observers.

Table 3: 201805 Number of observations by observer

Obs	NumObs	Name
AAP	2	A. Patrick Abbott
AAX	19	Alexandre Amorim
AJV	16	J. Alonso
ARAG	31	Gema Araujo
ASA	30	Salvador Aguirre
ATE	28	Teofilo Arranz Heras
BARH	14	Howard Barnes
BERJ	21	Jose Alberto Berdejo
BMF	23	Michael Boschat
BRAD	22	David Branchett
BRAF	26	Raffaello Braga
BROB	25	Robert Brown
BSAB	18	Santanu Basu
CHAG	31	German Morales Chavez
CIOA	16	Ioannis Chouinavas

Continued on next page

Table 3: 201805 Number of observations by observer

Obs	NumObs	Name
CKB	19	Brian Cudnik
CNT	9	Dean Chantiles
CVJ	17	Jose Carvajal
DEMF	11	Frank Dempsey
DJOB	15	Jorge del Rosario
DMIB	26	Michel Deconinck
DROB	11	Bob Dudley
DUBF	27	Franky Dubois
ERB	24	Bob Eramia
FERJ	19	Javier Ruiz Fernandez
FLET	27	Tom Fleming
FLF	21	Fredirico Luiz Funari
FTAA	21	Tadeusz Figiel
FUJK	19	K. Fujimori
HAYK	21	Kim Hay
HIVB	2	Ivan Hajdinjak
HMQ	6	Mark Harris
HOWR	22	Rodney Howe
JDAC	11	David Jackson
JGE	8	Gerardo Jimenez Lopez
JPG	6	Penko Jordanov
KAPJ	27	John Kaplan
KNJS	31	James & Shirley Knight
KROL	22	Larry Krozel
LEVM	22	Monty Leventhal
LRRA	9	Robert Little
MARE	12	Enrico Mariani
MCE	23	Etsuiku Mochizuki
MILJ	12	Jay Miller
MJAF	31	Juan Antonio Moreno Quesada
MJHA	27	John McCammon
MUDG	16	George Mudry
MWU	15	Walter Maluf
OATS	7	Susan Oatney
ONJ	21	John O'Neill
RLM	9	Mat Raymonde
SDOH	31	Solar Dynamics Obs - HMI
SIMC	13	Clyde Simpson
SMNA	3	Michael Stephanou
SNE	4	Neil Simmons
SONA	18	Andries Son
STAB	27	Brian Gordon-States
SUZM	21	Miyoshi Suzuki
TESD	31	David Teske

Continued on next page

Table 3: 201805 Number of observations by observer

Obs	NumObs	Name
TPJB	5	Patrick Thibault
URBP	8	Piotr Urbanski
VARG	31	A. Gonzalo Vargas
VIDD	12	Daniel Vidican
WCHD	5	Charles White
WILW	27	William M. Wilson
Totals	1184	65

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

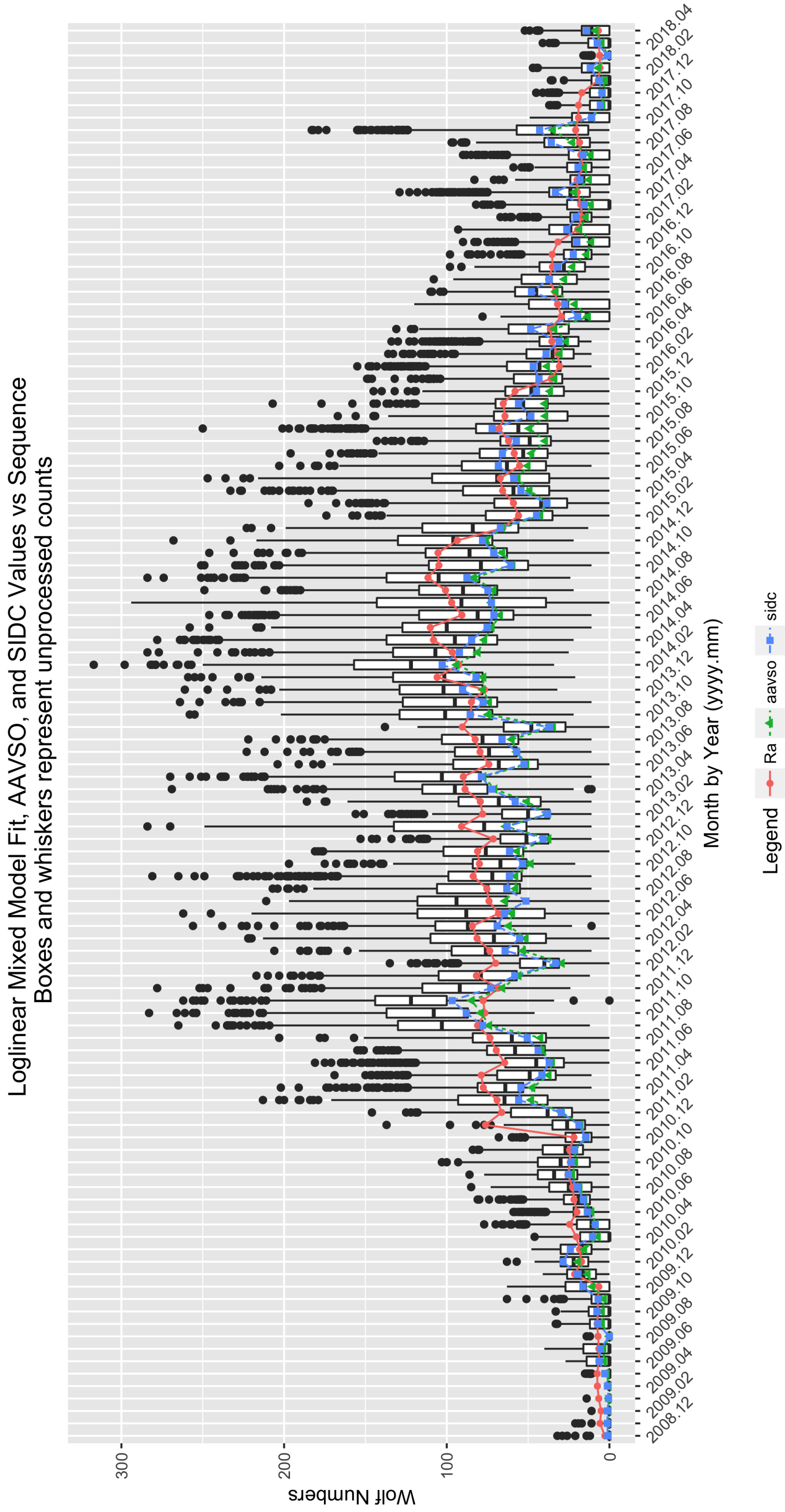

Figure 7: GLMM fitted data for R_a. AAVSO data: https://www.aavso.org/category/tags/solar-bulletin. SILSO data: WDC-SILSO, Royal Observatory of Belgium, Brussels