Solar Bulletin

THE AMERICAN ASSOCIATION OF VARIABLE STAR OBSERVERS
SOLAR SECTION

Rodney Howe, Kristine Larsen, Co-Chairs
c/o AAVSO, 185 Alewife Brook Parkway,
Cambridge, MA 02138 USA

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

Volume 78 Number 9 September 2022

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. The sudden ionospheric disturbance report is in Section 2. The relative sunspot numbers are in Section 3. Section 4 has endnotes.

1 Sunspot Group Numbers Compared

Submitted by Jamie Riggs, Ph.D.

As we saw last month, a common question in determining monthly Sunspot numbers is how similar are the counts across observers after accounting for seeing, equipment used, filter type, etc. The object of this article is to account for Sunspot group count consistency across observers. Last month, an exploratory data analysis (EDA) visually compared the group counts of selected observers who posted Sunspot counts to the AAVSO Solar Section database [AAVSO, 2022]. These selected observers submitted counts each month during the years spanning January 2000 to August 2022, which defines the study period.

Recall that the data for observers contributing Sunspot numbers over the study period were subsetted to a dozen observers. (More observers are expected to be added.) The data for these observers were screened to assure there was no missing data throughout the 22-year period. Smoothing was employed to impute missing group counts as needed. The EDA and resulting data conditioning prepared these data for modeling using the DeepAR recurrent neural network methodology implemented by Amazon, Inc. for multiple concurrent time series data. See the August 2022 Solar Bulletin for further details on the use of DeepAR (`https://docs.aws.amazon.com/sagemaker/latest/dg/deepar.html`) and recurrent neural networks.

The objective of using DeepAR on the 12 AAVSO observers' Sunspot group counts is to find forecasts for each observer based on each observer's group count history, including possible latent interplay among the observers. Last month, the report presented only a visual comparison. This article reports on a statistical comparison of the DeepAR forecasts. We reproduce Figure 1 from last month, which shows the individual historical group counts (black curves) with 48 months of forecasted (red curves) group counts, one plot per observer. Observers ARAG, BROB, BXZ, CKB, DUBF, FLET, HKY, KNJS, and RJV have visually similar group count histories, and thereby, forecasts as determined by the DeepAR algorithm. Observers BARH and CHAG have similar histories and forecasts. Observer SIMC is possibly unique among this group of 12 observers.

We use a nonparametric comparison of the observers' forecasts, as the data are not multivariate normal. See, for example, Conover [1999] and Hettmansperger and McKean [2011]. The test uses

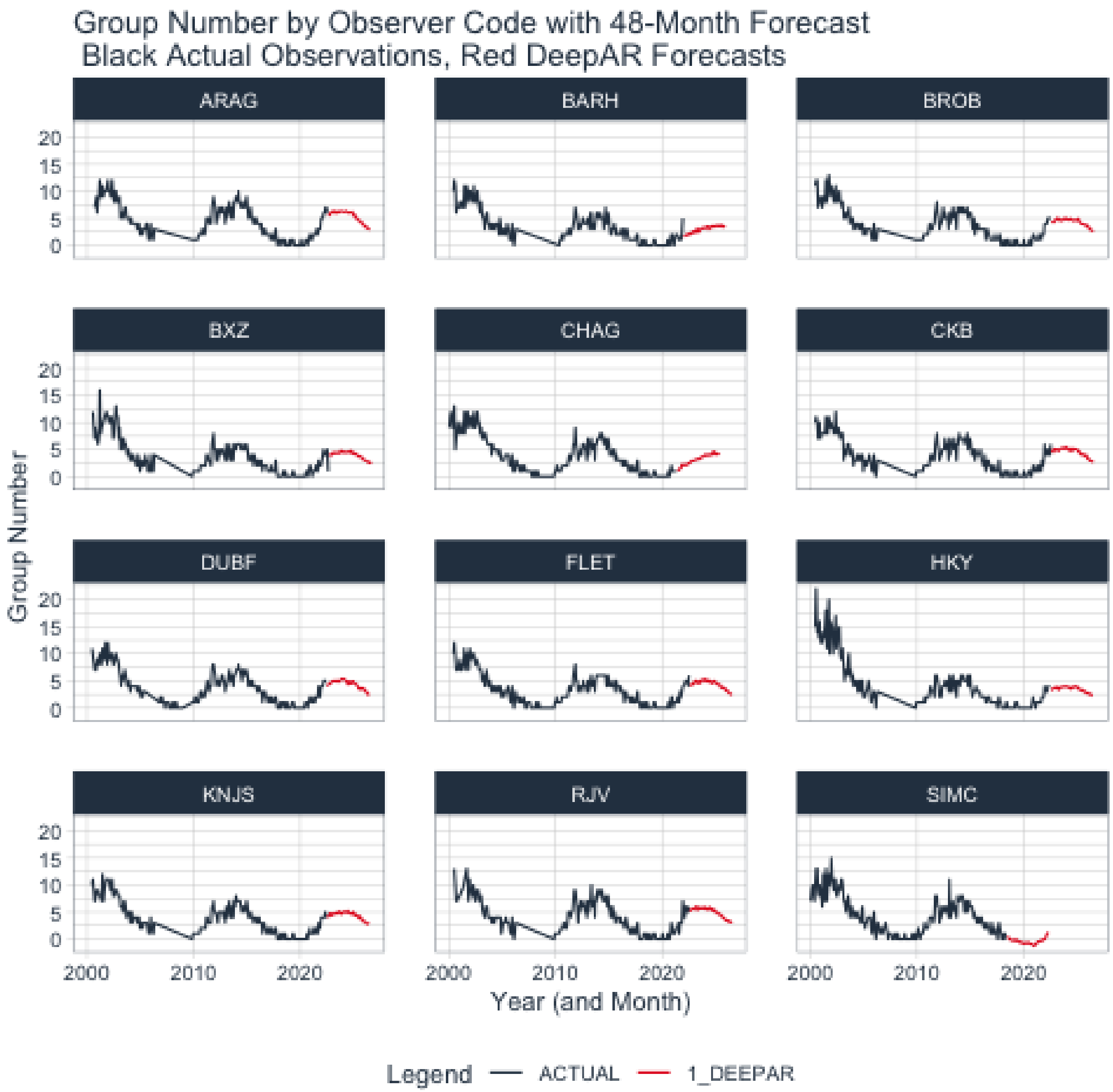

Figure 1: Twelve AAVSO observers' group counts. The black curves are the historical group counts (ACTUAL) from January 2000 to August 2022. The red curves (1_DEEPAR) are the observers' forecasts from September 2022 to August 2026.

F-approximations for the ANOVA Type test, the Wilks' Lambda Type test, the Lawley Hotelling Type test, and the Bartlett Nanda Pillai Type test statistics, as well as a permutation test for each. The tests compare the multivariate distributions of the different samples, and computes nonparametric relative effects. The test statistics are given in Table 1 below. In summary, there is **at least one** observer whose group counts are significantly different from the others.

Table 1: Table of nonparametric DeepAR forecast tests. Each test shows there is at least one observer who is statistically different from the others.

Test Type	Test Statistic	df1	df2	P-value	Permutation Test p-value
ANOVA type test	132.137	11.000	564	0	0
McKeon approx. for the Lawley Hotelling Test	132.137	11.000	564	0	0
Muller approx. for the Bartlett-Nanda-Pillai Test	131.907	11.019	564	0	0
Wilks Lambda	132.137	11.000	564	0	0

An examination of the Relative Effects in Table 2 suggests 8 observers have relative effects values greater than 0.4, thus having a large effect on the group count forecasts (Cohen's criteria has 0.1 as small, 0.25 as medium, and 0.4 as large). One observer has a small effect on group count forecasts with a relative effect size less than 0.1. The observers whose relative effect sizes are between 0.1 and 0.4 have a medium effect on group count forecasts.

Table 2: DeepAR forecast relative effects. The relative effects suggest 8 observers have relative effects values that are greater than 0.4, thus having a large effect on the group count forecasts (Cohen's criteria has 0.1 as small, 0.25 as medium, and 0.4 as large). One observer has a small effect on group count forecasts with a relative effect size less than 0.1. The observers whose relative effect sizes are between 0.1 and 0.4 have a medium effect on group count forecasts.

Observer	Relative Effect
ARAG	0.855540
BARH	0.221820
BROB	0.630430
BXZ	0.426180
CHAG	0.153750
CKB	0.677660
DUBF	0.693580
FLET	0.625000
HKY	0.364290
KNJS	0.574150
RJV	0.735930
SIMC	0.041667

The question as to which observers have group count forecasts that are statistically the same may be answered using the table of 95% confidence intervals (CI) about the median values of the forecasts. When any two observers' 95% CI overlap, then the group count median forecast values are statistically the same. It is clear from the CI table below (Table 3) that ARAG and SIMC do not share CI overlap with any of the other observers. All observers other than ARAG and SIMC have overlapping CI with at least one other observer.

Table 3: 95% Confidence Intervals (CI) with median values.

Statistic	ARAG	BARH	BROB	BXZ	CHAG	CKB	DUBF	FLET	HKY	KNJS	RJV	SIMC
uci	6.88	3.25	5.32	4.5	2.46	5.41	5.50	5.25	4.09	5.00	5.92	-0.03
med	6.58	2.96	5.07	4.37	2.19	5.22	5.25	5.06	3.94	4.79	5.66	-0.10
lci	6.28	2.67	4.83	4.17	1.92	5.04	5.01	4.87	3.78	4.58	5.39	-0.17

The box plot in Figure 2 ("Group Counts vs. Observer") shows the relationship of the observers to each other. The ordinate ("Group Count") ranges from 0 to 8 counts of group forecasts. The abscissa is categorized alphabetically by observer identification code. Above each box plot, each with its 95% CI about the median, is the upper 95% CI, the median value, and the lower 95% CI which is also in the 95% Confidence Intervals table (Table 3). The plots show that observer SIMC shows a clear difference from the others. The confidence interval comparison above shows that ARAG also is statistically different.

It is clear that the DeepAR model needs further tuning, as negative counts are among the forecasts. This is physically not possible, indicating additional work is needed. In addition, it will be interesting to add additional observers to increase the number of comparisons.

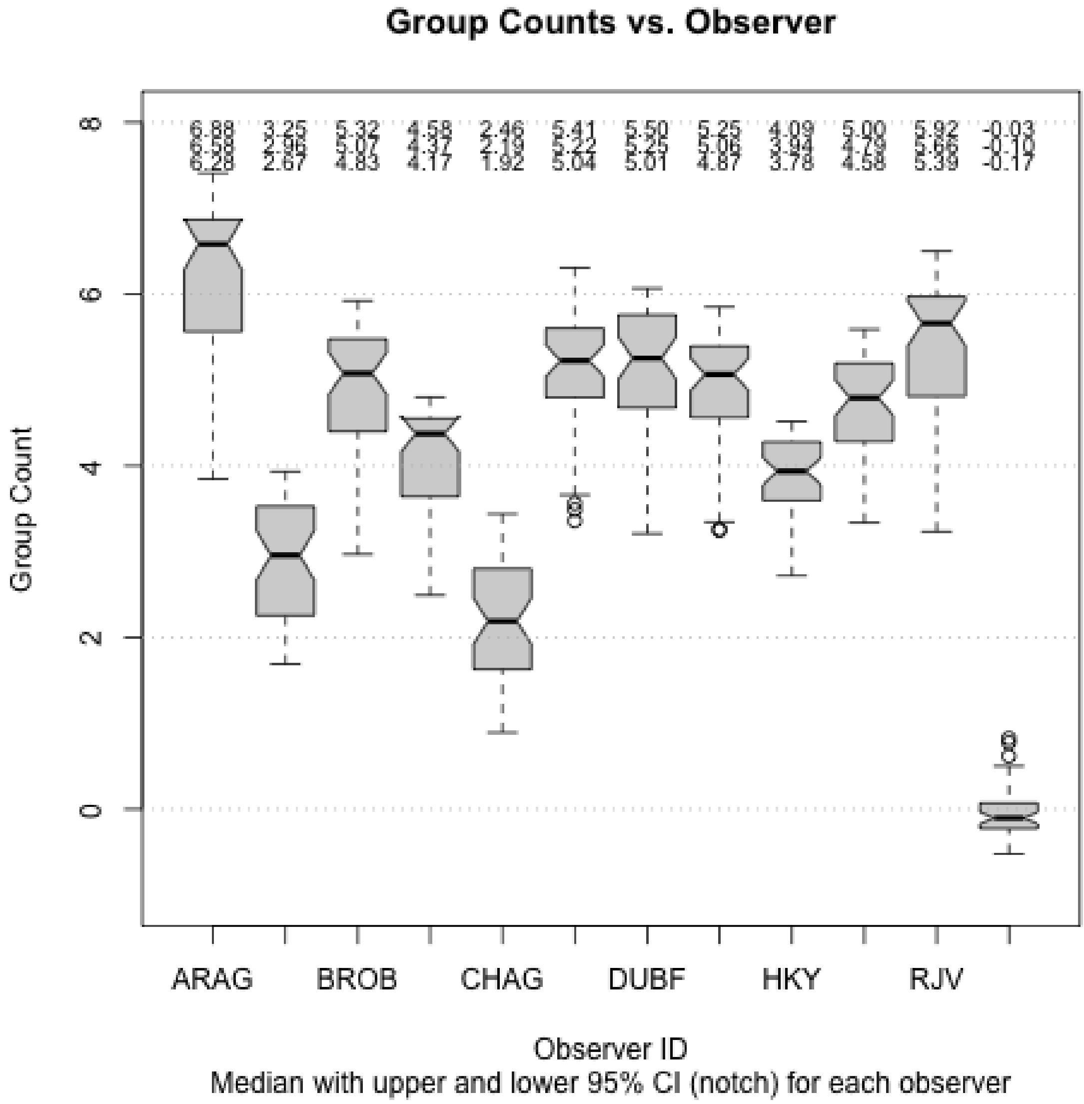

Figure 2: The relationship of the observers' Sunspot group counts with box plots. The ordinate ("Group Count") ranges from 0 to 8 counts of group forecasts. The abscissa is categorized alphabetically by observer identification code.

2 Sudden Ionospheric Disturbance (SID) Report

2.1 SID Records

September 2022 (Figure 3) there were a group of C- and M-class flares on the 30th of September recorded here in Fort Collins, Colorado.

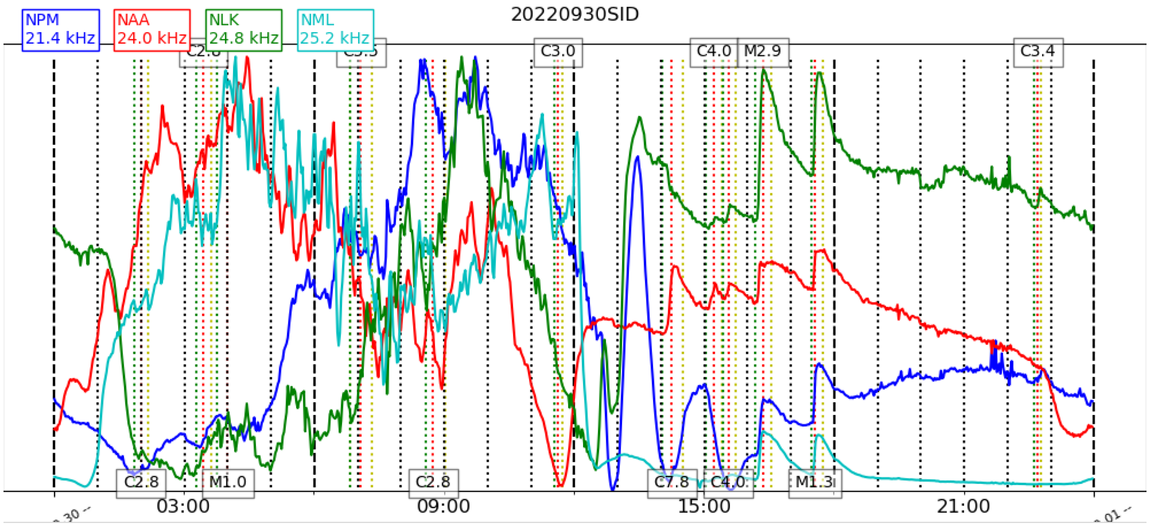

Figure 3: VLF recording on the 30th of September.

2.2 SID Observers

In September 2022, 15 AAVSO SID observers submitted VLF data as listed in Table 4.

Table 4: 202209 VLF Observers

Observer	Code	Stations
R Battaiola	A96	HWU
J Wallace	A97	NAA
L Loudet	A118	DHO
J Godet	A119	GBZ GQD ICV
F Adamson	A122	NWC
J Karlovsky	A131	DHO NAA TBB
R Mrllak	A136	GQD NSY
S Aguirre	A138	NPM NAA
G Silvis	A141	NAA NML NLK
K Menzies	A146	NAA
L Pina	A148	NAA NLK NML
J Wendler	A150	NAA
H Krumnow	A152	FTA GBZ HWU
J DeVries	A153	NLK
R Mazur	A155	NLK NML

Figure 4 depicts the importance rating of the solar events. The duration 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.

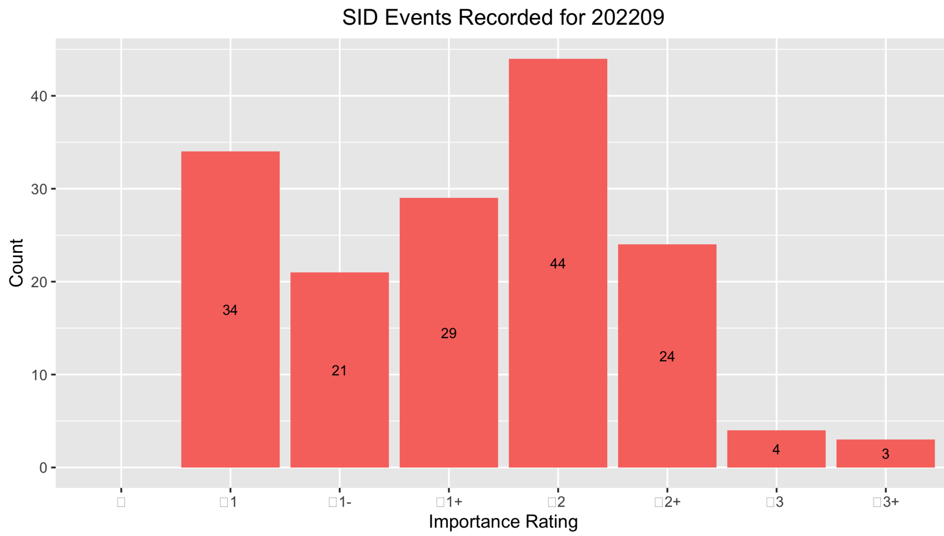

Figure 4: VLF SID Events.

2.3 Solar Flare Summary from GOES-16 Data

In September 2022, there were 317 GOES 16 XRA flares: 13 M-class, 293 C-class, 11 B-class; about the same flaring this month compared to last, although mostly in the C-class range (Figure 5).

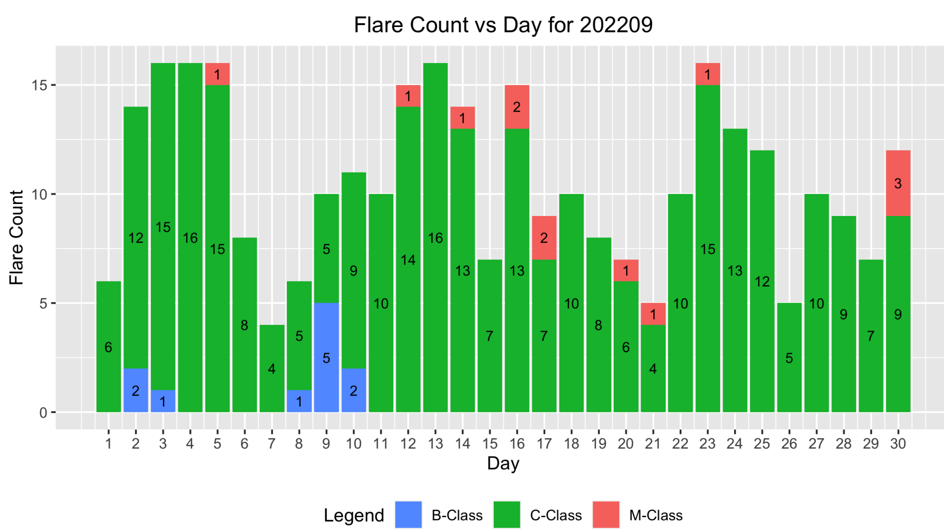

Figure 5: GOES-16 XRA (NOAA, 2022) flares.

3 Relative Sunspot Numbers (R_a)

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 Structured Query Language (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 September 2022. These counts are reported by the day of the month. 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.

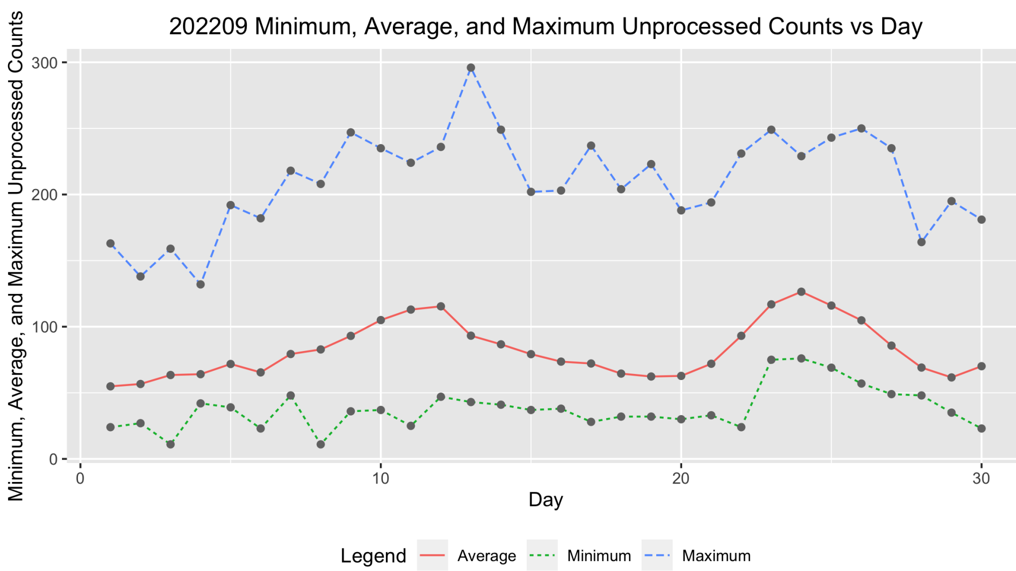

Figure 6: Raw Wolf number average, minimum and maximum by day of the month for all observers.

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 (`https://adsabs.harvard.edu/full/1949PASP...61...13S`). The Shapley method is a statistical model that agglomerates variation due to random effects, such as observer group selection, and fixed effects, such as seeing condition. The raw Wolf averages and calculated R_a are seen in Figure 7, and Table 5 shows the Day of the observation (column 1), the Number of Observers recording that day (column 2), the Raw Wolf number (column 3), and the Shapley Correction (R_a) (column 4).

Table 5: 202209 American Relative Sunspot Numbers ($\mathrm{R_a}$).

Day	Number of Observers	Raw	R_a
1	38	58	49
2	36	59	48
3	35	69	57
4	37	66	56
5	33	73	63
6	39	68	55
7	35	84	70
8	34	90	75
9	35	101	82
10	36	114	92
11	40	115	95
12	39	117	98
13	28	97	81
14	30	95	78
15	37	83	68
16	41	81	66
17	43	75	63
18	38	68	57
19	32	64	52
20	37	67	55
21	37	78	64
22	35	103	83
23	35	118	105
24	33	127	115
25	35	124	103
26	36	108	92
27	39	90	75
28	33	71	60
29	37	67	56
30	34	76	64
Averages	35.9	86.9	72.6

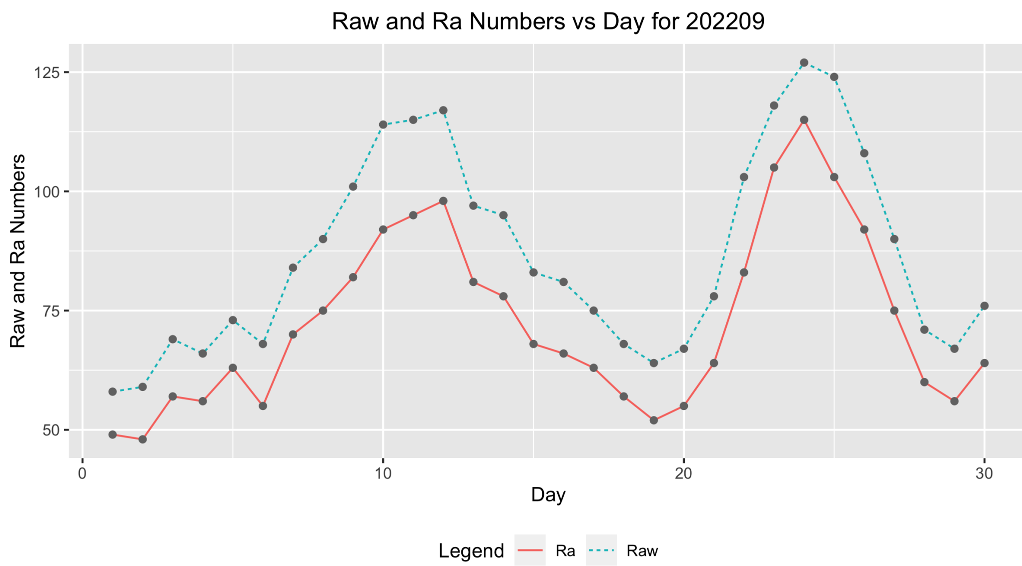

Figure 7: Raw Wolf average and R_a numbers by day of the month for all observers.

3.3 Sunspot Observers

Table 6 lists the Observer Code (column 1), the Number of Observations (column 2) submitted for September 2022, and the Observer Name (column 3). The final row gives the total number of observers who submitted sunspot counts (66), and total number of observations submitted (1077).

Table 6: 202209 Number of observations by observer.

Observer Code	Number of Observations	Observer Name
AAX	17	Alexandre Amorim
AJV	16	J. Alonso
ARAG	29	Gema Araujo
ASA	18	Salvador Aguirre
ATE	23	Teofilo Arranz Heras
BATR	7	Roberto Battaiola
BMF	23	Michael Boschat
BMIG	21	Michel Besson
BROB	21	Robert Brown
BXZ	26	Jose Alberto Berdejo
BZX	24	A. Gonzalo Vargas
CANG	2	Andrew Corkill
CIOA	3	Ioannis Chouinavas
CKB	20	Brian Cudnik
CNT	27	Dean Chantiles
CVJ	8	Jose Carvajal
DARB	16	Aritra Das
DELS	10	Susan Delaney

Continued

Table 6: 202209 Number of observations by observer.

Observer Code	Number of Observations	Observer Name
DFR	9	Frank Dempsey
DMIB	30	Michel Deconinck
DUBF	26	Franky Dubois
EHOA	16	Howard Eskildsen
ERB	24	Bob Eramia
FERA	15	Eric Fabrigat
FLET	28	Tom Fleming
GIGA	22	Igor Grageda Mendez
HALB	15	Brian Halls
HKY	20	Kim Hay
HMQ	5	Mark Harris
HOWR	22	Rodney Howe
HRUT	21	Timothy Hrutkay
IEWA	25	Ernest W. Iverson
ILUB	3	Luigi Iapichino
JDAC	7	David Jackson
JGE	4	Gerardo Jimenez Lopez
JSI	6	Simon Jenner
KAND	28	Kandilli Observatory
KAPJ	23	John Kaplan
KNJS	27	James & Shirley Knight
LEVM	5	Monty Leventhal
LJAE	1	Jay Lavender
LKR	12	Kristine Larsen
MARC	3	Arnaud Mengus
MARE	12	Enrico Mariani
MCE	20	Etsuiku Mochizuki
MJHA	21	John McCammon
MLL	10	Jay Miller
MMI	30	Michael Moeller
MWU	15	Walter Maluf
OAAA	26	Al Sadeem Astronomy Obs.
ONJ	12	John O'Neill
RJUB	6	Justus Randolph
RJV	20	Javier Ruiz Fernandez
SDOH	30	Solar Dynamics Obs - HMI
SNE	11	Neil Simmons
SQN	2	Lance Shaw
SRIE	18	Rick St. Hilaire
TDE	29	David Teske
TNIA	7	Nick Tonkin
TPJB	4	Patrick Thibault
TST	25	Steven Toothman

Continued

Table 6: 202209 Number of observations by observer.

Observer Code	Number of Observations	Observer Name
URBP	24	Piotr Urbanski
VIDD	16	Dan Vidican
WGI	1	Guido Wollenhaupt
WND	4	Denis Wallian
WWM	26	William M. Wilson
Totals	1077	66

3.4 Generalized Linear Model of Sunspot Numbers

Dr. Jamie Riggs, Northwestern University and Thomas Jefferson University, 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. For more details: *A Generalized Linear Mixed Model for Enumerated Sunspots* (see 'GLMM06' in the sunspot counts research page at http://www.spesi.org/?page_id=65).

Figure 8 shows the monthly GLMM R_a numbers for a rolling eleven-year (132-month) window beginning within the 24th solar cycle and ending with last month's sunspot numbers. 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 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 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.

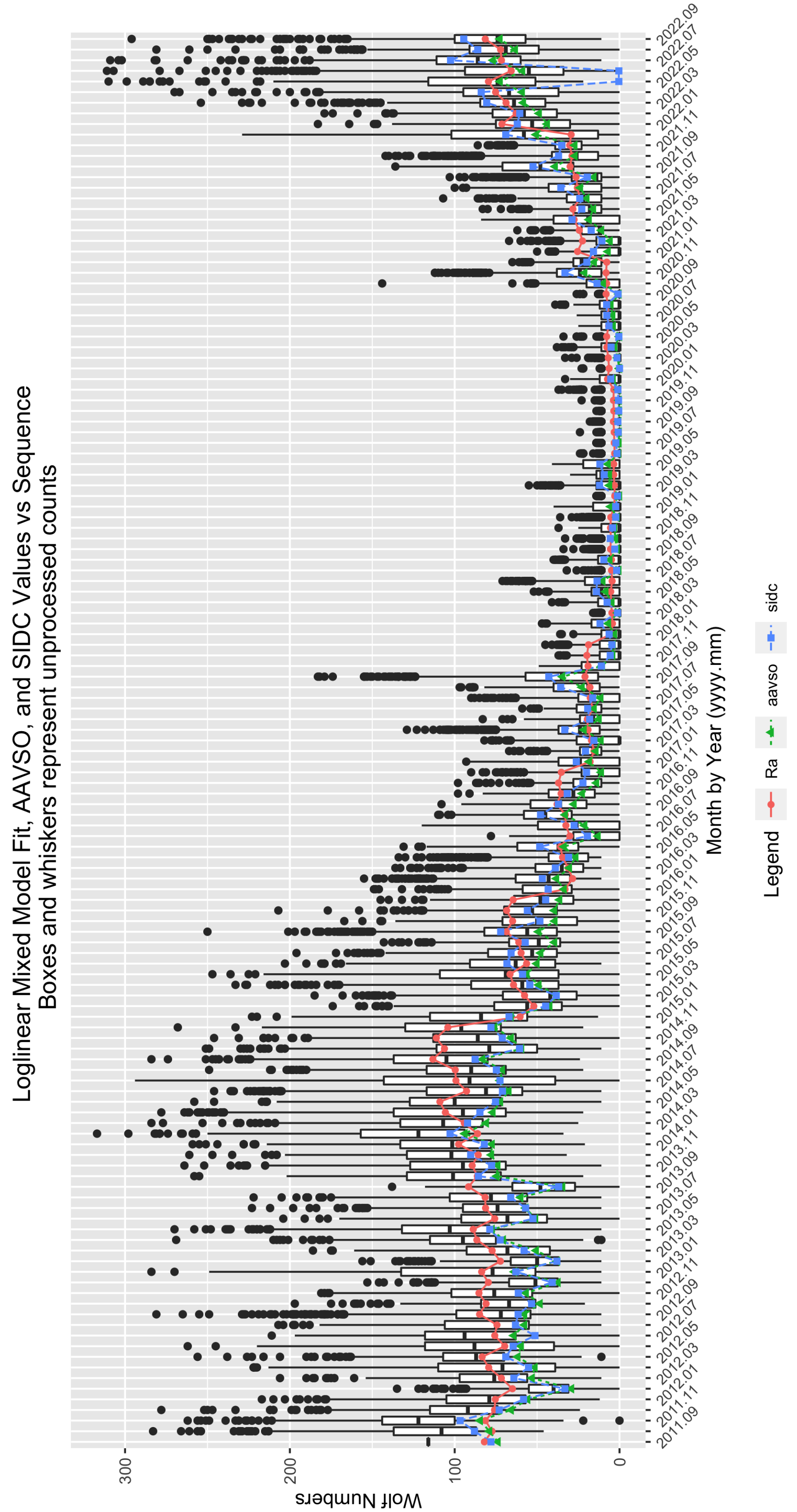

Figure 8: GLMM fitted data for R_a. AAVSO data: https://www.aavso.org/category/tags/solar-bulletin. SIDC data: WDC-SILSO, Royal Observatory of Belgium, Brussels

4 Endnotes

- Sunspot Reports: Kim Hay solar@aavso.org
- SID Solar Flare Reports: Rodney Howe rhowe137@gmail.com

Software

The following are the R and R packages used in comparing Sunspot group counts.
R [R Core Team, 2021]: A language and environment for statistical computing.
R package modeltime.gluonts Dancho [2022]: modeltime.gluonts: GluonTS Deep Learning_. R package version 0.3.1, `https://github.com/business-science/modeltime.gluonts`
R package Tidymodels Kuhn and Wickham [2020]: Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. `https://www.tidymodels.org` **R package Tidyverse** Wickham et al. [2019]: A collection of R functions.
R package timetk Dancho and Vaughan [2022]: C Tool Kit for Working with Time Series in R_. R package version 2.8.1, `https://CRAN.R-project.org/package=timetk`

References

AAVSO. Aavso solar observing section, 2022. URL `https://www.aavso.org/solar`.

W. J. Conover. *Practical Nonparametric Statistics.* John Wiley & Sons, Inc, New York, 3rd edition, 1999.

Matt Dancho. *modeltime.gluonts: 'GluonTS' Deep Learning*, 2022. URL `https://github.com/business-science/modeltime.gluonts`. R package version 0.3.1.

Matt Dancho and Davis Vaughan. *timetk: A Tool Kit for Working with Time Series in R*, 2022. URL `https://CRAN.R-project.org/package=timetk`. R package version 2.8.1.

T.P. Hettmansperger and J.W. McKean. *Robust Nonparametric Statistical Methods.* Monographs on Statistics and Applied Probability 119. Taylor and Francis Group, LLC, Boca Raton, Florida, 2011.

Max Kuhn and Hadley Wickham. *Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles.*, 2020. URL `https://www.tidymodels.org`.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2021. URL `https://www.R-project.org/`.

Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4 (43):1686, 2019. doi: 10.21105/joss.01686.

U.S. Dept. of Commerce, NOAA, Space Weather Prediction Center, 2022, (ftp://ftp.swpc.noaa.gov/pub/indices/events/)

SILSO, World Data Center - Sunspot Number and Long-term Solar Observations. (2022). Sunspot number catalogue, 1850-2022 [data set]. Royal Observatory of Belgium. (https://www.sidc.be/silso/datafiles)