Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Discussion on "Poisson-FOCuS: An Efficient Online Method for Detecting Count Bursts with Application to Gamma Ray Burst Detection" by Ward, Dilillo, Eckley and Fearnhead

Yang Chen

To cite this article: Yang Chen (2025) Discussion on "Poisson-FOCuS: An Efficient Online Method for Detecting Count Bursts with Application to Gamma Ray Burst Detection" by Ward, Dilillo, Eckley and Fearnhead, Journal of the American Statistical Association, 120:549, 26-30, DOI: 10.1080/01621459.2024.2402961

To link to this article: https://doi.org/10.1080/01621459.2024.2402961

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2025, VOL. 120, NO. 549, 26–30: Applications and Case Studies Discussion
https://doi.org/10.1080/01621459.2024.2402961

Discussion on "Poisson-FOCuS: An Efficient Online Method for Detecting Count Bursts with Application to Gamma Ray Burst Detection" by Ward, Dilillo, Eckley and Fearnhead

Yang Chen

Department of Statistics, University of Michigan, Ann Arbor, MI

1. Introduction

It is my great pleasure to discuss "Poisson-FOCuS: An Efficient Online Method for Detecting Count Bursts with Application to Gamma Ray Burst Detection" by Ward, Dilillo, Eckley, and Fearnhead (Ward et al. 2024). This article proposes efficient computational algorithms for detecting Gamma Ray Bursts (GRB), which are targeted for use by satellites for real-time detection. The article extends the FOCuS (Romano et al. 2023) approach for online change detection to Poisson counts. In this discussion, I will focus on the major merits that I learned from this article, the challenges of the current algorithm in practice, and potential applications of the algorithm to solar eruption detection (after proper adaptations). These merits and challenges are recognized on the basis of the GRB presented by the authors, and more so in my own research experience working with physical phenomena such as solar flares, solar energetic particles, and geomagnetic storms.

Here, we take solar flares as a motivating example for the following discussions. Solar flares are eruptions of energy from the Sun that can last from minutes to hours. Predicting solar flare events is traditionally based on physics models, but nowadays, it is showing promise with data-driven approaches. See Chen et al. (2024) for a review of the challenges and opportunities of solar imaging data for statisticians. There are several features of the solar flare data that, in my opinion, resemble those of GRBs and thus could benefit from the FOCuS and/or Poisson-FOCuS method.

As published by the Space Weather Prediction Center news on Monday, July 29, 2024, 03:09 UTC (SWPC/NOAA 2024b) (the same week this discussion was written), a significant X1.5 (R3-Strong) solar flare was observed at 29/0233 UTC from SWPC Region 3764 (S05W04). The immediate effect of this event is that "users of high-frequency (HF) signals may experience temporary degradation or complete loss of signal on much of the sunlit side of the Earth." As published by follow-up news on Monday, July 29, by the SWPC/NOAA (SWPC/NOAA 2024c), "Geomagnetic storm watches are out for 29–31 July due to a number of coronal mass ejections (CMEs). Solar activity was elevated through the weekend and various events, including solar flares and filament eruptions, were associated with CMEs." The CMEs typically follow strong solar flares and can send charged particles to space. If the Earth is in the path of a CME, the charged particles can disrupt or fail satellites in orbit and expose high-flying airplanes to radiation. The SWPC monitors solar eruptions, including flares, CMEs, and geomagnetic storms, and gives forecasts in real-time.

In Figure 1, we show the GOES X-ray flux from July 24 to July 30, where peak times of the detected flares of M class are labeled with blue vertical dashed lines, and the detected X class flare on July 29 is labeled with a red vertical dashed line on the top panel. In the bottom panel of Figure 1, we zoomed in to July 29 and labeled the start and end times of the X class flare with pink dashed lines and the peak time with a red solid line. Note that X-class flares are the most powerful solar flares, which correspond to peak flux crossing the 10^{-4} threshold as shown in Figure 1, and are quite rare to observe. M-class flares are 10 times less intense than X-class flares. This year, 2024, we have observed quite a few X-class flares since the Sun is approaching the solar maximum.

2. Merits

The paper is an impressive application and case study due to several aspects.

2.1. Low Computational Cost for Real-time Application

First and most importantly, Ward et al. (2024) explicitly focuses on the computational cost, giving a step-by-step detailed comparison of computational efficiency with competing approaches that scientists currently adopt. This is one of the things that caught my attention strongly since this is not a "standard" comparison that everyone does for either theoretical or applied statistics papers. I have worked a lot on developing novel statistical methods for scientific data, especially in the field of physical sciences in the past. Computational efficiency is one of the most critical aspects of the method that scientists use to pick up and adopt it. In some of my own experiences, rigorous statistical models that take care of data distributions do not outperform off-the-shelf deep learning/ neural network approaches in terms of computational efficiency (e.g., Sun et al. 2023). However, we do gain accuracy by modeling dependency struc-

CONTACT Yang Chen ychenang@umich.edu Department of Statistics, University of Michigan, Ann Arbor, MI.

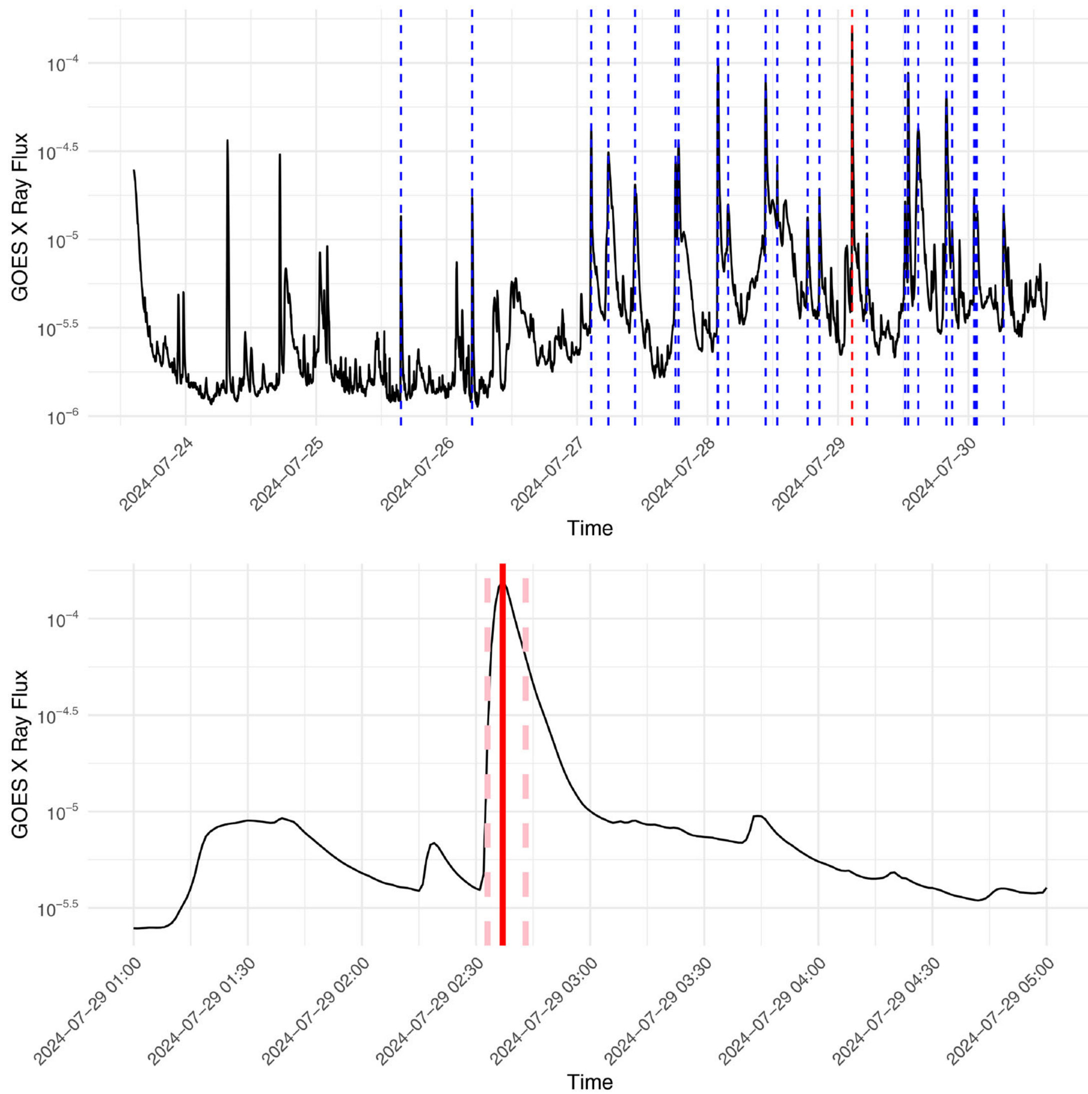

Figure 1. A visualization of the GOEX X-ray flux data from July 24 to 30, 2024. The data is downloaded from the Space Weather Prediction Center (SWPC) of the National Oceanic and Atmospheric Administration (NOAA, SWPC/NOAA (2024a)) on July 30, 2024. The black lines correspond to the 1-min cadence data of the GOES X-ray flux, and the blue vertical lines correspond to the M-class flares on the top panel. On the bottom panel, the pink vertical dashed lines correspond to the start and end time of the X class flare on July 29 (shown by the only red dashed vertical line on the top panel), which peaks at the time of the red vertical solid line.

tures (Marshall et al. 2021; Sun et al. 2022), taking care of homogeneity and heterogeneity structures (Chen et al. 2016; Trangucci et al. 2023; Viet Do et al. 2024; Iong et al. 2024a, 2024b), and providing uncertainty estimates of physically meaningful parameters (Chen et al. 2019) and predicted quantities (Sun and Chen 2024). This article hits me as a reminder to further strengthen my work in real-time forecasting of solar eruptions (see, e.g., Chen et al. 2024, for a review) and geomagnetic perturbations (Iong et al. 2022, 2024a) in terms of computational cost and implementation efficiency, which will ultimately make statistical prediction models more competitive in operational settings.

2.2. Connections with and Improvements upon Window Methods

Window methods are classical for change point detection algorithms but face challenges in operational settings. Ward et al. (2024) not only compares with the window methods but also draws connections and points out improvements of Poisson-FOCuS upon window methods. The visualizations given in Figure 7 of Ward et al. (2024) are highly illustrative and straightforward to understand, which represents the kind of figures I hope to see in many papers introducing new methodology and to incorporate in my future paper writing, too.

2.3. Capability of Detecting both Bright and Dim GRBs Online

The GRBs are rare and come in various intensities and durations, which is the same for major space weather events such as solar eruptions and geomagnetic storms. The detection of strong and weak signals with one algorithm that trades off the false positives and false negatives is nontrivial. In Ward et al. (2024), the functional pruning gives the signal strength a continuous treatment. This is through tracking the functional curves of the test statistics as a function of signal strength μ and updating only

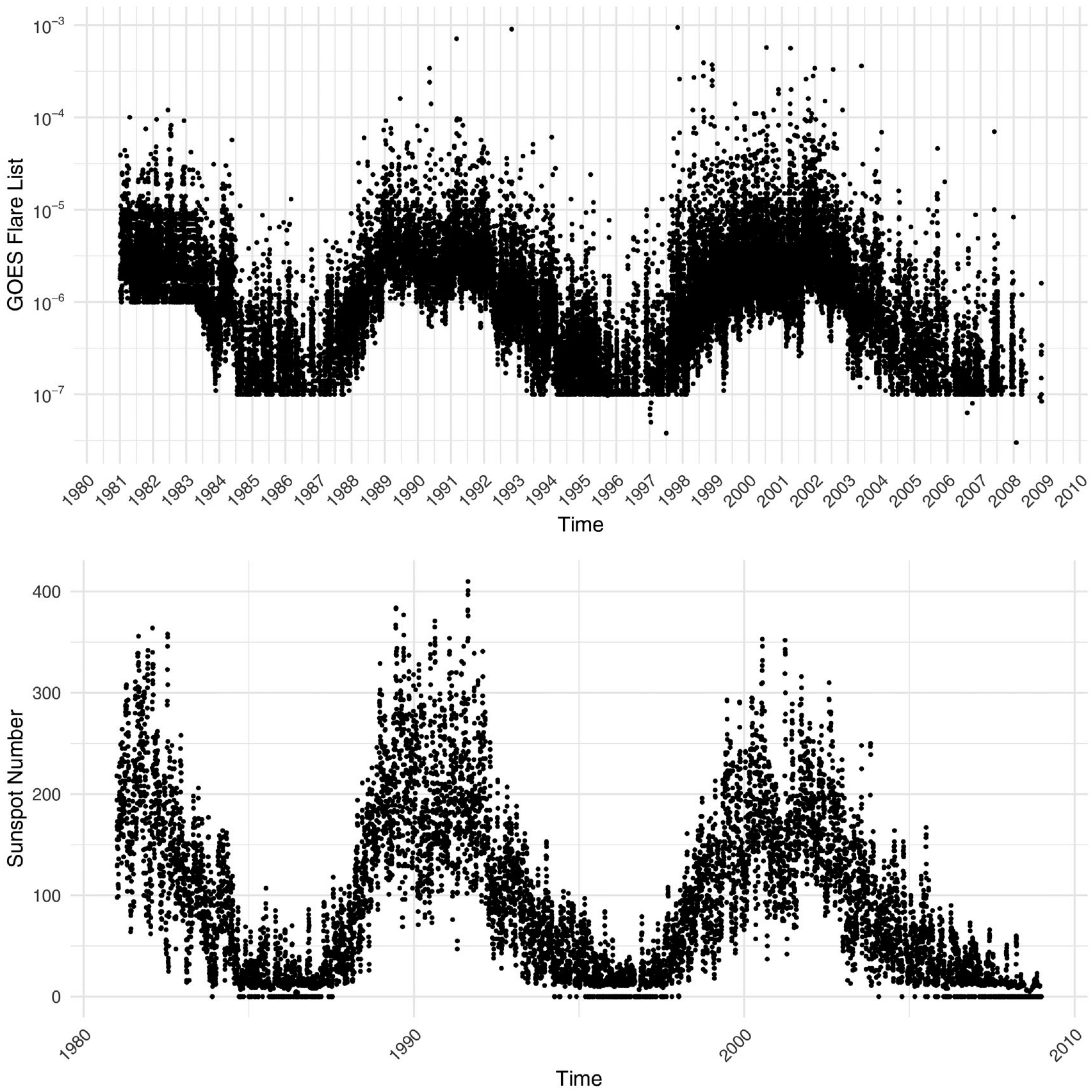

Figure 2. The solar flare peak intensities from the GOES flare list (top panel) and the sunspot numbers (bottom panel) from 1980 to 2010.

the coefficients in a sequential manner as new data comes in. Due to the temporal nature of observations and the sequential nature of the algorithm, the duration of an event is thus automatically taken care of. The online nature of the algorithm and the low false positive rate make it highly practical in operational settings. This property is highly desirable in practice, especially considering the solar flare detection shown in Figure 1 and the fact that solar flares come at varying magnitudes (A/B/C/M/X classes that differ from each other by a multiplier of 10 in peak intensities).

3. Challenges and applications

3.1. Model Assumptions on Signal Intensity

The main modeling assumption I would like to see future work address is the constant signal intensity assumption, which does not capture the shape of a GRB. In the typical solar flare setting, the signal often comes suddenly but decays gradually (see the second panel in Figure 1), with some exceptions. A parametric model for the flaring signal will make the likelihood ratio statistic not as simple as it currently stands. Still, better capture of the signal start and end times is definitely needed, which currently can only be modified/estimated after post-processing. It remains unclear at which time points the current algorithm with the assumed constant signal strength detects/defines the gradually increasing phase and the likely exponential decay phase of a solar flare event. In fact, the Bayesian block algorithm proposed in Scargle (1998) and improved in Scargle et al. (2013) takes into account the varying signal shapes. Furthermore, the alternative hypothesis with Poisson rate given by $\mu\lambda$ where λ is the background rate means the signal strength is proportional to the background level, which is slightly counter-intuitive in scientific settings where signal strength is typically assumed to be independent of the background level, which is *added* toward the signal level. For example, additive background models are assumed in image deconvolution in astronomy (Jones et al. 2015; Donath et al. 2024) and in X-ray spectra analysis (Kaastra 2017). This is an arguable assumption when detecting strong solar flares (of M or X class). The intensities of detected solar flares do follow the same trend as the "background level", that is, general solar activity level. In other words, in a highly volatile period during which many weaker (A/B/C class) flares happen (corresponding to a higher background value in Ward et al. (2024)), empirically, we can see that there is a higher chance of strong flares with higher peak intensity being recorded. This is exemplified by Figure 2, which shows that the GOES flare list has peak flare intensities following the same trend as the sunspot numbers (solar cycle progression). However, the weaker flares that are not recorded in solar maximum years can provide useful information for flare forecasting model training if they can be detected. In summary, the assumptions are quite stringent as they currently stand in Ward et al. (2024) and need further validation or modification prior to being adopted for other applications such as solar flare detection.

3.2. Background Estimation and Deduction

It is pointed out in Ward et al. (2024) that robust methods need to be adopted for the background estimation, the underestimation of which can result in false detections. An exponential smoother is applied to FERMI data. This is based on the assumption that background is typically assumed to be much smoother and varies at a much lower rate than the signals. In addition, I wonder if there can be a real-time feedback loop for signal detection and background estimation. The detected signals can be verified in the "sanity checking" stages. If, during a certain amount of time, we observe false detection as an anomaly over a long period of time, then a decision can be made on re-estimating the background rate. This refinement step resembles a conventional procedure of iteratively updating background and signal strength, see, for example, Zhang et al. (2023) for a data-driven background correction application in astronomy.

3.3. Explicit Studies on False Positive and Negative Rates

As mentioned in the Introduction of Ward et al. (2024), a good detection system should have a very low false positive rate. Ward et al. (2024) adopts the notion of "k-sigma event" as a measure of Type-I error rate in sequential testing. I am very curious about the performances under different levels of k regarding false positive and false negative rates. The false positives in GRB detection result in unnecessary computational costs for more complex sanity checking, while the false negatives in GRB detection simply mean missing an event. Neither of these two scenarios is desired in practice; thus, it will be valuable to know explicitly the potential "risks" of false positive and false negative events before applying the algorithm in operation.

3.4. Applications to Solar Flare Detection

The solar eruption detection algorithms are rather primitive, just like in the GRB setting, and primarily are based on window methods and thresholding. Machol, Codrescu, and Peck (2024) describes the detailed flare detection pipeline and algorithms adopted in detail. In the solar flare detection problem, the false negatives and false positives need to be balanced carefully, especially in the real-time setting, since they both have undesirable consequences. I would like to adapt the FOCus and/or Poisson-FOCuS ideas to develop an online flare detection algorithm with the science-verified GOEX X-ray flux data shown in Figure 1.

3.5. Comparison with Bayesian Block Algorithms

Scargle (1998) proposes a detection algorithm that outputs the most probable segmentation of the observation into time intervals during which the photon arrival rate is perceptibly constant. The performance is demonstrated using the BATSE γ-ray burst data. Later, Scargle et al. (2013) improves the method by finding the optimal segmentation of the data in the observation interval. The method can be used in either a real-time trigger mode or a retrospective mode. They also extended the method to account for piecewise linear and piecewise exponential signals. It will be great to see the authors' comments on this thread of literature and some comparisons of results in various realistic settings.

4. Conclusion

In summary, I learned a lot from Ward et al. (2024) and got motivated to develop computationally efficient solar eruption detection algorithms that can be applied in real time. Again, I want to congratulate the authors on this excellent piece of work that sends a strong message of revolutionary statistical methods for cutting-edge scientific problems.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

The author gratefully acknowledges fundings from NSF and NASA. YC is supported by NSF DMS 2113397, NSF PHY 2027555, NASA 22-SWXC22_2-0005, and 22-SWXC22_2-0015.

ORCID

Yang Chen http://orcid.org/0000-0002-9516-8134

References

Chen, Y., Shen, K., Shan, S.-O., and Kou, S. (2016), "Analyzing Single-Molecule Protein Transportation Experiments via Hierarchical Hidden Markov Models," *Journal of the American Statistical Association*, 111, 951–966. [27]

Chen, Y., Meng, X.-L., Wang, X., van Dyk, D. A., Marshall, H. L., and Kashyap, V. L. (2019), "Calibration Concordance for Astronomical Instruments via Multiplicative Shrinkage," *Journal of the American Statistical Association*, 114, 1018–1037. [27]

Chen, Y., Manchester, W., Jin, M., and Pevtsov, A. (2024), "Solar Imaging Data Analytics: A Selective Overview of Challenges and Opportunities," arXiv preprint arXiv:2405.12331v2. [26,27]

Donath, A., Siemiginowska, A., Kashyap, V. L., van Dyk, D. A., and Burke, D. (2024), "Joint Deconvolution of Astronomical Images in the Presence of Poisson Noise," arXiv preprint arXiv:2403.13933. [28]

Iong, D., Chen, Y., Toth, G., Zou, S., Pulkkinen, T., Ren, J., Camporeale, E., and Gombosi, T. (2022), "New Findings From Explainable Sym-h Forecasting Using Gradient Boosting Machines," *Space Weather*, 20, e2021SW002928. [27]

Iong, D., McAnear, M., Qu, Y., Zou, S., and Chen, G. T. Y. (2024a), "Sparse Variational Cntaminated Noise Gaussian Process Regression for Forecasting Geomagnetic Perturbations," *Data Science in Science*, 3, 1–17. [27]

Iong, D., Zhao, Q., and Chen, Y. (2024b), "A Latent Mixture Model for Heterogeneous Causal Mechanisms in Mendelian Randomization," *The Annals of Applied Statistics*, 18, 966–990. [27]

Jones, D. E., Kashyap, V. L., and Van Dyk, D. A. (2015), "Disentangling Overlapping Astronomical Sources Using Spatial and Spectral Information," *The Astrophysical Journal*, 808, 137. [28]

Kaastra, J. S. (2017), "On the Use of C-Stat in Testing Models for X-ray Spectra," *Astronomy & Astrophysics*, 605, A51. [28]

Machol, J., Codrescu, S., and Peck, C. (2024), "User's Guide for Goes-R XRS l2 Products," available at *https://data.ngdc.noaa.gov/platforms/solar-spaceobserving-satellites/goes/goes16/l2/docs/GOES-R_XRS_L2_Data_Users_Guide.pdf*. [29]

Marshall, H. L., Chen, Y., Drake, J. J., Guainazzi, M., Kashyap, V. L., Meng, X.-L., Plucinsky, P. P., Ratzlaff, P., van Dyk, D. A., and Wang, X. (2021), "Concordance: In-Flight Calibration of X-ray Telescopes Without Absolute References," *The Astronomical Journal*, 162, 254. [27]

Romano, G., Eckley, I. A., Fearnhead, P., and Rigaill, G. (2023), "Fast Online Changepoint Detection via Functional Pruning Cusum Statistics," *Journal of Machine Learning Research*, 24, 1–36. [26]

Scargle, J. D. (1998), "Studies in Astronomical Time Series Analysis. V. Bayesian Blocks, a New Method to Analyze Structure in Photon Counting Data," *The Astrophysical Journal*, 504, 405. [28,29]

Scargle, J. D., Norris, J. P., Jackson, B., and Chiang, J. (2013), "Studies in Astronomical Time Series Analysis. VI. Bayesian Block Representations," *The Astrophysical Journal*, 764, 167. [28,29]

Sun, H., and Chen, Y. (2024), "Conformalized Tensor Completion with Riemannian Optimization," arXiv preprint arXiv:2405.00581. [27]

Sun, H., Hua, Z., Ren, J., Zou, S. Sun, Y., and Chen, Y. (2022), "Matrix Completion Methods for the Total Electron Content Video Reconstruction," *The Annals of Applied Statistics*, 16, 1333–1358. [27]

Sun, H., Manchester, W., Jin, M., Liu, Y., and Chen, Y. (2023), "Tensor Gaussian Process with Contraction for Multi-Channel Imaging Analysis," in *Proceedings of the Fortieth International Conference on Machine Learning (ICML)*, (Vol. 202), pp. 32913–32935. [26]

SWPC/NOAA. (2024a), "GOES X-Ray Flux," available at *https://www.swpc.noaa.gov/products/goes-x-ray-flux*. Online; accessed 30 July 2024. [27]

——— (2024b), "Strong (R3) Flare Activity Observed - 29 JULY 2024," available at *https://www.swpc.noaa.gov/news/strong-r3-flare-activity-observed-29-july-2024*. Online; accessed 30 July 2024. [26]

——— (2024c), "Geomagnetic Storm Watched in Effect 29-31 JULY (UP TO G3; Strong)," available at *https://www.swpc.noaa.gov/news/geomagnetic-storm-watches-effect-29-31-july-g3-strong*. Online; accessed 30 July 2024. [26]

Trangucci, R., Chen, Y., and Zelner, J. (2023), "Modeling Racial/Ethnic Differences in Covid-19 Incidence with Covariates Subject to Nonrandom Missingness," *The Annals of Applied Statistics*, 17, 2723–2758. [27]

Viet Do, B., Chen, Y., Nguyen, L., and Manchester, W. (2024), "Uncovering Heterogeneity of Solar Flare Mechanism with Mixture Models," *Frontiers in Astronomy and Space Sciences*, 11, 1229092. [27]

Ward, K., Dilillo, G., Eckley, I., and Fearnhead, P. (2024), "Poisson-Focus: An Efficient Online Method for Detecting Count Bursts with Application to Gamma Ray Burst Detection," *Journal of the American Statistical Association*, the issue, DOI:10.1080/01621459.2023.2235059. [26,27,29]

Zhang, X., Algeri, S., Kashyap, V., and Karovska, M. (2023), "A Novel Approach to Detect Line Emission Under High Background in High-Resolution X-ray Spectra," *Monthly Notices of the Royal Astronomical Society*, 521, 969–983. [29]