Flare Finder: Detecting Near-Ultraviolet Flares in M-dwarfs in Swift UVOT Event Data

Sophia Wang
Institute for Computing in Research
July 2024

Abstract

This paper introduces the algorithms and usage of Flare Finder, a program developed to detect near-ultraviolet (NUV) flares on M-dwarfs from data from the Neil Gehrels Swift Observatory (*Swift*). In previous studies, flares on M-dwarfs are first imaged optically, and telescopes encompassing other wavelengths with narrower ranges are turned to record data accordingly; however, NUV flares on M-dwarfs do not always produce an optical flare distinguishable from noise, biasing the set of observed flares. Given event data from the *Swift* ultraviolet and optical telescope (UVOT), Flare Finder solves the issue of bias by identifying candidates for flare stars and timestamps of flare(s) without optical input. The program bins individual star data into light curves, and signal-to-noise ratios (SNRs) are calculated for each flare duration window of time for each star. The program is tested using *Swift* UVOT event data of a recent flare on flare star YZ Canis Minoris (YZ CMi) and its results, successes, and possible improvements are discussed.

1 Introduction

Representing about 70% of all stellar objects, M-dwarfs are characterized by their low mass and long lifespans (Bochanski et al., 2010). Observations of M-dwarfs have revealed that most are favorable hosts for large systems of exoplanets, including rocky- and Jupiter-sized planets; roughly one-third of rocky planets orbit in the circumstellar habitable zones of M-dwarfs (Shields et al., 2016). M-dwarfs also flare more frequently compared to other stars, as they generate stronger magnetic fields (Shulyak, D. et al., 2019). Flares occur when stored magnetic energy is converted into kinetic energy and released, heating up the star and causing it to emit radiation observable across the electromagnetic spectrum (Kowalski, 2024).

Understanding and characterizing stellar flares on M-dwarfs is critical for determining the habitability of orbiting exoplanets, especially near-ultraviolet (NUV) radiation. Giant stellar flares create high levels of UV radiation that alter planetary atmospheres, decreasing habitability (Howard et al., 2020). On the other hand, certain levels of ultraviolet radiation are essential for prebiotic chemistry, such as ribonucleotides (Ranjan et al., 2017).

Typically, NUV wavelength data of M-dwarf flares is collected through targeted investigation. Flares are first observed through the optical wavelength using wide-field cameras, for example the Transiting Exoplanet Survey Satellite (TESS); telescopes with smaller fields measuring in different wavelengths are then pointed directly at the identified target, for example the Neil Gehrels Swift Observa-

tory (*Swift*) ultraviolet and optical telescope (UVOT). However, this method of flare detection biases observations against detecting high-energy NUV flares, and only 28% of NUV flares have an accompanying optical flare, as the flare is diluted in noise in the light curve (Paudel et al., 2024). Thus, relying on optical data to discover flare stars is insufficient to answer the question of habitability with regard to ultraviolet light.

This paper introduces a method to programmatically detect flares in *Swift* UVOT event data to fill the gap in detected flares without optical input, and is organized as follows. Section 2 outlines details of the *Swift* UVOT, event data, and flare detection algorithm. The algorithm is run on example data with a known flare occurrence, and the results are discussed in Section 3. Section 4 gives a summary of the main results from this project and future improvements.

2 Data and Algorithm

2.1 The *Swift* UVOT and Event Data

The three-telescope *Swift* observatory was launched in 2004 and remains in service today. One of these telescopes is the UVOT, which counts photons in the optical and ultraviolet wavelengths (1600 to 6000 Å) with a timing resolution of approximately 11 ms (NASA, 2014). Due to its limited field of view (17 by 17 arcminutes), to this point, it is intentionally pointed at a known flare star to detect NUV flare data.

The UVOT saves data in three modes, one of which is Event mode. Event mode saves a list of observed photons, or events, and their timestamps, respective coordinates in three different formats, exposure references, and quality flags (Immler et al., 2008). For the Flare Finder algorithm, the data of interest is the timestamp, X and Y position, and quality flag of each event.

2.2 Flare Finder Algorithm

The Flare Finder algorithm first cleans data for quality and contiguity. Events with a nonzero quality flag are removed, as well as events at timestamps with few other observations. The remaining data is processed for star detection, then flare detection. The program was first written using Jupyter notebook software and later converted to a single Python file (Kluyver et al., 2016).

2.2.1 Star Detection

During star detection, data is first binned along the X and Y into a 1000 by 1000 2D histogram. Then, synthetic aperture photometry is used to identify individual stars, using a similar but simplified application of the DAOPHOT algorithm (Stetson, 1987). A circular aperture spanning 2 by 2 pixels is used as the window for photon counts of a star, and an annulus aperture with an outer radius of 11 pixels and an inner radius of roughly 9 pixels is used to calculate the local background for that star.

For every 2 by 2 pixel area on the 2D histogram, the total star counts (C_t) in the circular aperture are added together, and the background counts (C_s) are similarly summed. Moreover, the area of the signal and sky apertures are calculated ($A_t = 4$ and $A_s = 184$, respectively). The signal in the aperture N_t is calculated with equation 1.

$$N_t = C_t - A_t \times \frac{C_s}{A_s} \qquad (1)$$

The sky background per pixel N_s is also calculated as:

$$N_s = \frac{A_t \times C_s}{A_s} \qquad (2)$$

Using output from equations 1 and 2, the total noise σ is thus calculated as:

$$N = \sqrt{N_t + N_s} \qquad (3)$$

2

The final signal-to-noise ratio (SNR) is, naturally, calculated with equation 4.

$$SNR = \frac{N_t}{N} \quad (4)$$

After calculation of all SNRs in the image, Flare Finder identifies the mean $\bar{x}$ and standard deviation σ of the ratios, which by definition are approximately 0 and 1, respectively. Stars, which are bright relative to their local background, by definition have a larger SNR compared to the mean; the algorithm applies a simple threshold of five standard deviations from the mean to isolate stars from the background. That is, for any point i, j, if $SNR_{i,j} > \bar{x} + 5 \times \sigma$, then that point belongs to a star in the original data.

Following this threshold application, a basic floodfill algorithm is used to determine all points of individual stars. The coordinates and average SNR belonging to each star are saved, and the program continues on to flare detection.

2.2.2 Flare Detection

The algorithm for flare detection uses a nested iterative process to search for flares in every star.

Flare Finding Algorithm

  for *star* in *stardata* **do**
    Get star data
    for *time window length* **do**
      for *timestamp* in *star data* **do**
        $SNR \leftarrow$ calculated SNR ratio
        return SNR
      end for
      $Outliers \leftarrow$ list of outlier SNRs
      return $Outliers$
    end for
  end for

All in all, there are three for loops to iterate over the star, the length of the time window, and the timestamp in the star data. For each star, the raw data from the UVOT event list is gathered, sorted, and binned into an Astropy TimeSeries with a bin size of one second. That is, all timestamps within the same second will be summed into one bin containing the total number of photon events for that second. The program then iterates over different signal window sizes to accommodate for flares of different lengths, and calculates the SNR for each window of time in the TimeSeries.

The SNR calculation over the time domain is relatively similar to the calculation for images, but with some noticeable changes. Rather than an annulus and circle, there is a signal window of length L_s, and two background windows to the left and right of that window, with lengths $L_{bg} = 1.5 \times L_s$.

$$N_b g = \frac{C_{bg}}{L_{bg}} \quad (5)$$

Equation 5 describes the mean counts per unit time N_{bg} in the background, where C_{bg} is the total photon counts in the background.

$$N_s = \frac{C_s}{L_s} - N_{bg} \quad (6)$$

Meanwhile, equation 6 is the calculation for the mean photon counts in the signal window, subtracting background noise, where C_s is the total photon counts in the signal window.

The total noise rate is also calculated using the standard deviation of the events in the background. For each bin i in the background, letting the number of photon events per second in that bin be x_i, then the total noise was calculated with equation 7.

$$N = \sqrt{\frac{\sum (x_i - N_{bg})^2}{L_{bg}}} \quad (7)$$

The SNR is still calculated using the same equation as 4. With N_s still as the mean signal rate, the SNR for a particular time window in a particular star is:

$$SNR = \frac{N_s}{N} \quad (8)$$

3

Like when detecting stars, the algorithm returns all the timestamps with an extremely high SNR (i.e. $SNR \geq \bar{x} + 5 \times \sigma$). The returned timestamps are all candidates of when a flare may have occurred.

3 Results

The Flare Finder algorithm was tested on recently procured *Swift* UVOT event data, including a flare from star YZ Canis Minoris (YZ CMi). The 2D histogram of the event data is shown below (Figure 1).
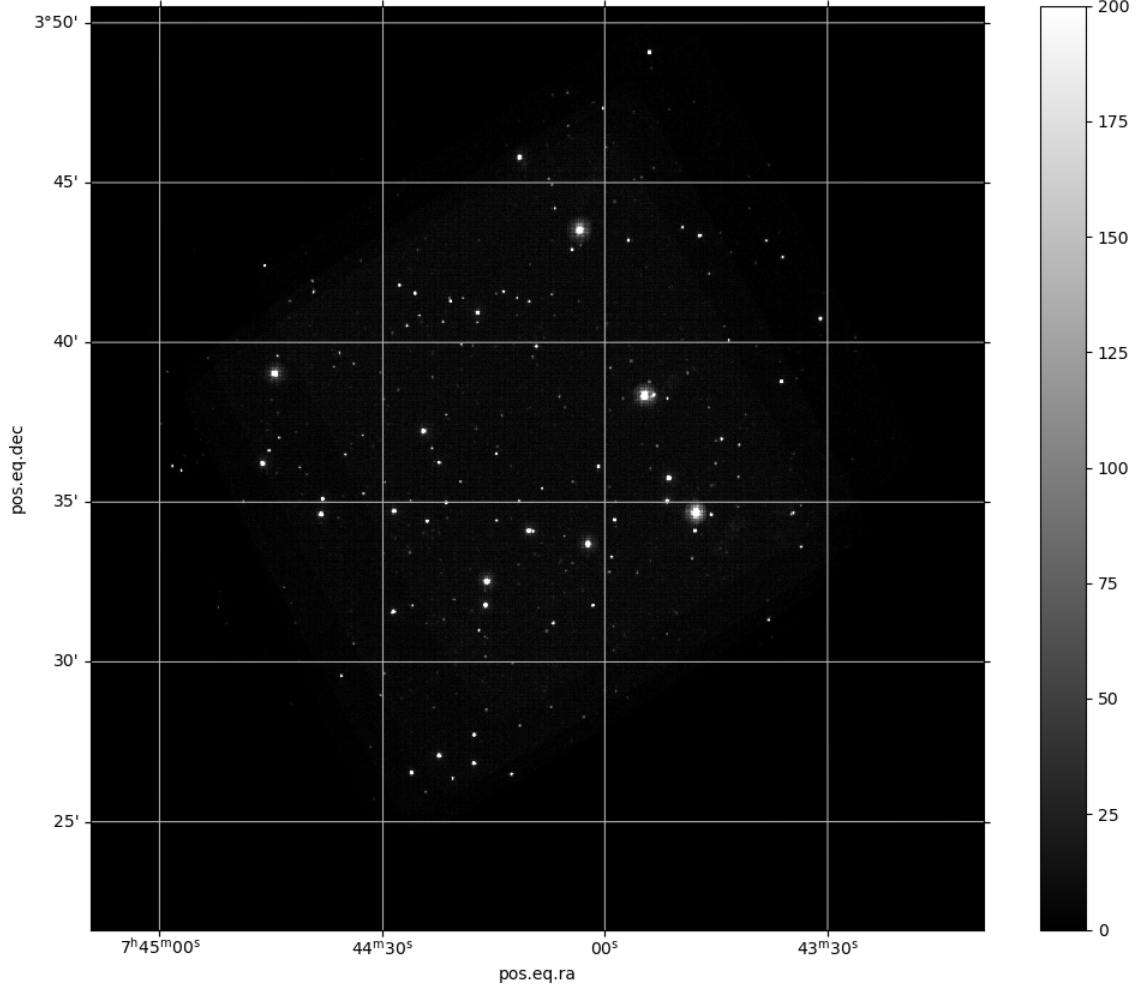

Figure 1: Grayscale 2D histogram of *Swift* UVOT event data.

Using the star detection algorithm detailed in Section 2.2.1, the SNRs of each pixel in Figure 1 was calculated with the annulus in Figure 2. The detected stars were plotted in Figure 3 on the left, while all points above the 5-standard-deviation threshold were plotted on the right.
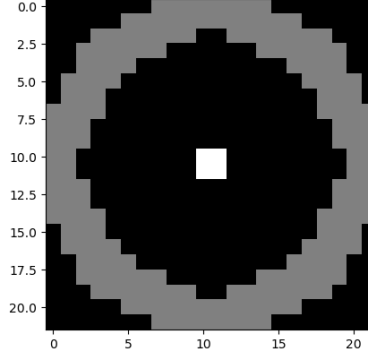
Figure 2: Annulus/circle aperture used for star detection, where the white section is the circle aperture and the gray annulus is the background aperture.
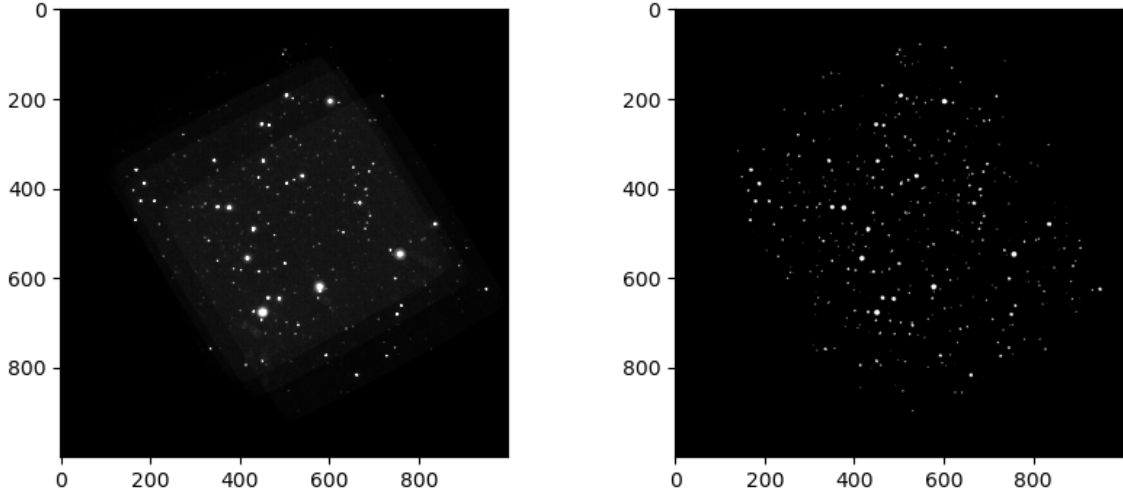

Figure 3: Star data after SNR ratios are calculated. Left: Visualization of SNR ratios of each point on the image. Right: Visualization of all points with SNRs above 5-standard-deviation threshold.

Noticeably, as seen in Figure 3, on the left, the calculated SNRs of the background are nonzero. Due to limb darkening, the brighter region of each star on the left is larger than the threshold-cut points on the right.

To test the Flare Finding algorithm on individual stars, two relatively large and bright stars were selected and analyzed for flares. From this point, the star in Figure 4 will be referred to as star A, and the star in Figure 5 will be referred to as star B.

For both stars, histograms of calculated SNR ratios for time windows of 10 seconds were also generated.
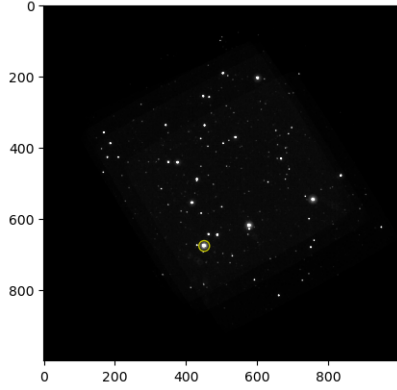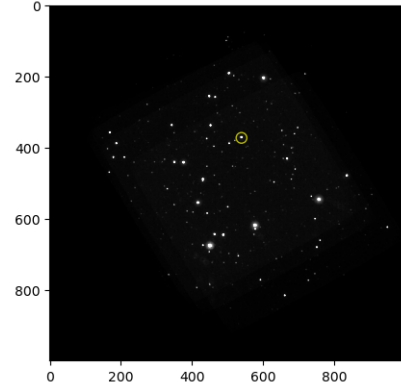

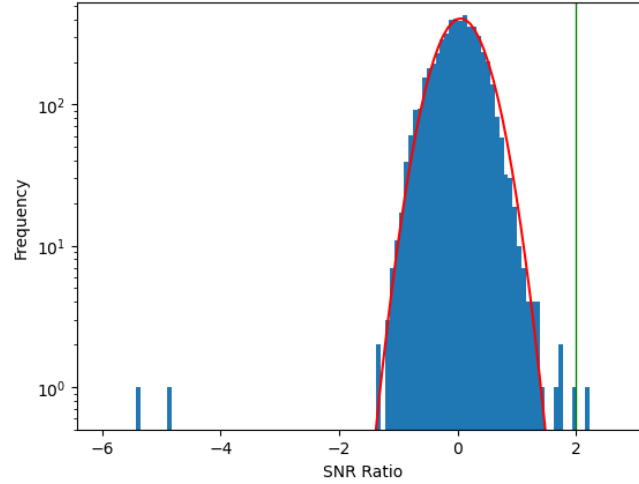Figure 4: Star A.

Figure 5: Star B.

Figure 6: Histogram of SNR ratios for Star A, with fitted Gaussian curve in red, and vertical bar five standard deviations from the mean in green.
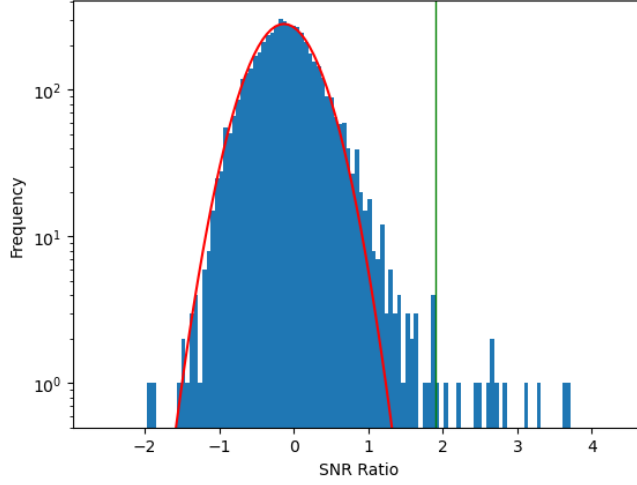
Figure 7: Histogram of SNR ratios for Star B, with red Gaussian curve and green vertical bar for 5-standard-deviation threshold.

Noticeably in Figure 6, there are few candidates for flares in a window of 10 seconds for star B. There are also two low outliers with SNRs between -4 and -6 that indicate a large dip in the signal window compared to peaks in the background; i.e. there are more photons detected in the background than in the signal window. Meanwhile in Figure 7, there are numerous timestamp candidates for flares and no low outliers.

It is important to note that number of outlier ratios in the histogram does not equal the number of probable flares in the data. Because the actual flare duration may be slightly shorter or longer than the signal window,

in this case 10 seconds, multiple consecutive timestamps may be marked as having a flare.

Figure 8 shows the timestamp with the largest calculated SNR of about 2.2310. Zooming out in Figure 9, it is clear that while this is a local peak, its height is still relatively similar to the extended background. Moreover, there are two spikes in the TimeSeries where the photon count approaches 60 counts/second, but because the duration of the peak is so short, for a window of 10 seconds, it becomes averaged into the background. Hence, if the signal window length is a poor estimate of the true flare duration, much shorter or longer flares may not be detected at all.
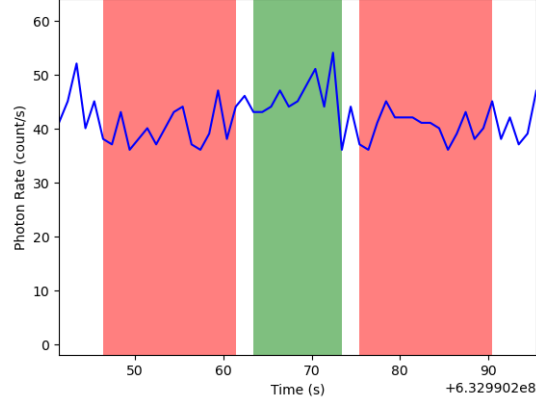
7

Figure 8: Relevant TimeSeries visualization for the timestamp with the largest calculated SNR for Star A, where the shaded green region represents the signal window, the two shaded red regions represent the background windows, and the blue line indicates the photon count per second.
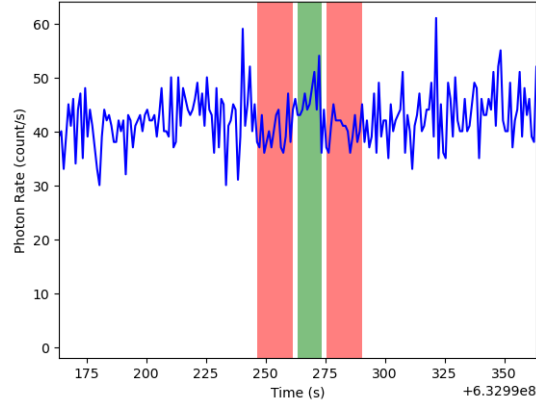

Figure 9: TimeSeries visualization of the peak in Figure 8, zoomed out so that there are about 80 seconds before and after the background windows.

The largest calculated SNR value for Star B was approximately 3.7143, indicating a very strong signal from the signal window. As seen in Figure 10, there is a dramatic increase at time 633000886, and the maximum photon rate of over 50 photons/second is much larger than the average in the local background mean rate of roughly 15.6333, indicative of a flare. Zooming out in Figure 11 reveals the entirety of the flare, which was largely undetected as the time window of 10 seconds was to short to detect the whole flare. As expected, this is a flare on M-dwarf flare star YZ CMi detected by Paudel et al. using TESS (2024).
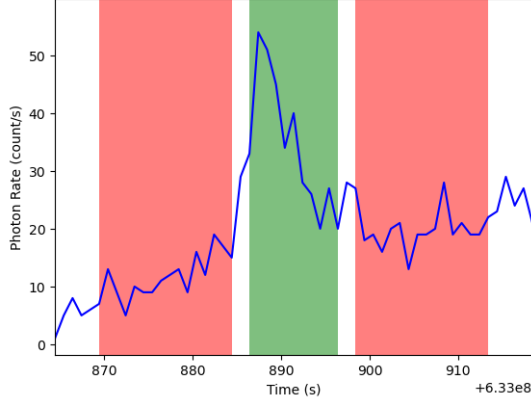
8

Figure 10: Relevant TimeSeries visualization for the timestamp with the largest calculated SNR for Star B, with a very large spike in the central green signal window.
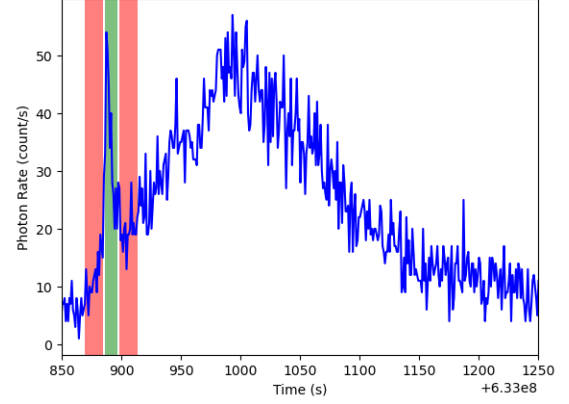

Figure 11: TimeSeries visualization of the detected peak in Figure 8, x-axis adjusted to show most of the flare ($t = 633000850$ to 633001200).
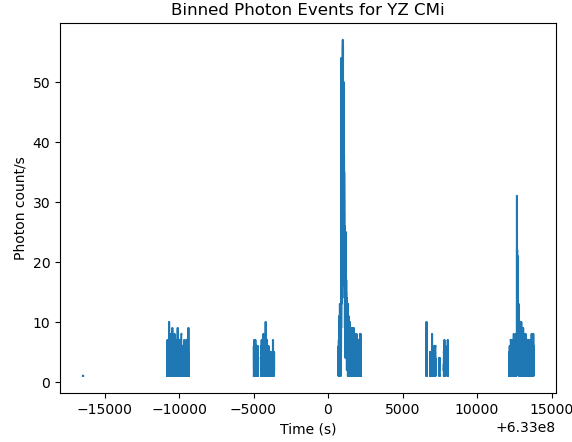

Figure 12: TimeSeries of entire binned photon data for Star B (YZ CMi) without SNR analysis.

Particularly large flares are potentially visible by eye, without using SNR analysis in the time domain (Figure 12). It is also easy to identify another between roughly $t = 633012000$ and 633013000, pictured in Figure 13. However, identifying flares by eye is time- and power-consuming as each star must be individually analyzed, for potentially many ter-abytes of data. This reveals the appeal of Flare Finder, as the entire process is automated from beginning to end, with flare candidates and visualizations directly reported and saved locally. For both observed spikes in photons/second in Figure 12, those timestamps are returned without guesswork or estimation.
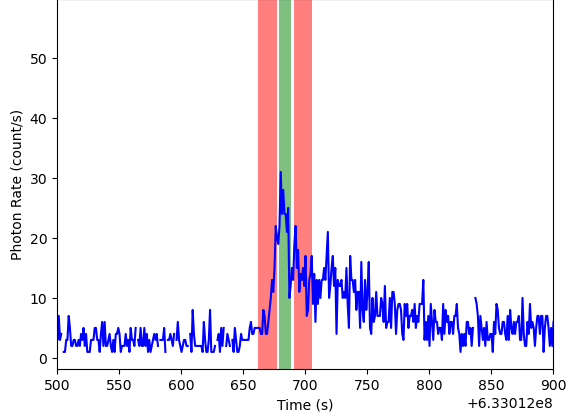
9

Figure 13: TimeSeries of the other, smaller, observed peak in photon event data for YZ CMi.

4 Summary and Possible Improvements

To reduce bias in NUV data for observed flares in M-dwarfs, Flare Finder is an algorithm designed to blindly search for flares and output flare star candidates in *Swift* UVOT event data. It iterates in polynomial time over any inputted UVOT event data and calculates SNRs in the time domain to determine timestamps of significantly higher photon output. When analyzing data from the *Swift* UVOT with flares on YZ CMi, the algorithm accurately reported the timestamps of both and provided other visualizations, including histograms of the SNRs.

Several small but persistent issues remain at hand, including noise distinction, the time window heuristic, and the SNR calculation over the time domain.

- **Noise distinction.** The current version of Flare Finder does not deal with high spikes of noise, as seen (potentially) in Figure 9, and could mistake them as very short flares.

- **Time window heuristic.** Flare Finder relies on an estimation for the length of the flare (e.g. 10 seconds), or sev-

eral estimations (5 *and* 10 seconds) to decrease processing time. This causes longer- or shorter-lasting flares to be ignored (e.g. in Figures 9 and 11). Optimizing the calculation process, or potentially using machine learning to choose better lengths for the time window, will increase Flare Finder's accuracy.

- **SNR calculation.** By definition, the standard deviation of the distribution of SNRs for each star should be 1. However, when analyzing the data, it consistently measured a standard deviation of about 0.5, indicating something in the calculation process may be systematically affecting the SNR.

5 Acknowledgements

I would like to specially thank my mentor, Aaron Tohuvavohu, for dedicating invaluable time, insights, and guidance to me for this project. I also extend thanks to the Institute for Computing in Research, especially Dr. Mark Galassi and Mr. José Medina-Hernandez, for offering me this opportunity to undertake this project.

Software: Astropy (Astropy Collaboration

et al., 2013; 2018; 2022), Jupyter Notebook (Kluyver et al., 2016), Numpy (Harris et al., 2020), Pillow (Clark, 2015), Scipy (Virtanen et al., 2020).

References

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., Lim, P. L., Crawford, S. M., Conseil, S., Shupe, D. L., Craig, M. W., Dencheva, N., Ginsburg, A., Vand erPlas, J. T., Bradley, L. D., Pérez-Suárez, D., de Val-Borro, M., Aldcroft, T. L., Cruz, K. L., Robitaille, T. P., Tollerud, E. J., . . . Astropy Contributors. (2018). The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package., *156*(3), Article 123, 123. https://doi.org/10.3847/1538-3881/aabc4f

Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., Earl, N., Starkman, N., Bradley, L., Shupe, D. L., Patil, A. A., Corrales, L., Brasseur, C. E., N"othe, M., Donath, A., Tollerud, E., Morris, B. M., Ginsburg, A., Vaher, E., Weaver, B. A., Tocknell, J., Jamieson, W., . . . Astropy Project Contributors. (2022). The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package., *935*(2), Article 167, 167. https://doi.org/10.3847/1538-4357/ac7c74

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A. M., Kerzendorf, W. E., Conley, A., Crighton, N., Barbary, K., Muna, D., Ferguson, H., Grollier, F., Parikh, M. M., Nair, P. H., . . . Streicher, O. (2013). Astropy: A community Python package for astronomy., *558*, Article A33, A33. https://doi.org/10.1051/0004-6361/201322068

Bochanski, J. J., Hawley, S. L., Covey, K. R., West, A. A., Reid, I. N., Golimowski, D. A., & Ivezić, Ž. (2010). The luminosity and mass functions of low-mass stars in the galactic disk. *The Astronomical Journal*, *139*(6), 2679–2699. https://doi.org/http://dx.doi.org/10.1088/0004-6256/139/6/2679

Clark, A. (2015). Pillow (pil fork) documentation. https://buildmedia.readthedocs.org/media/pdf/pillow/latest/pillow.pdf

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., . . . Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Howard, W. S., Corbett, H., Law, N. M., Ratzloff, J. K., Galliher, N., L. Glazier, A., Gonzalez, R., Soto, A. V., Fors, O., del Ser, D., & Haislip, J. (2020). Evryflare. iii. temperature evolution and habitability impacts of dozens of superflares observed simultaneously by evryscope and tess. *The Astrophysical Journal*, *902*(2), 115. https://doi.org/10.3847/1538-4357/abb5b4

Immler, S., Still, M., Boyd, P., Hollard, S., Marshall, F., Landsman, W., Rohrbach, G., Tripicco, M., & Wiegand, R. (2008, March).

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., & development team, J. (2016). Jupyter notebooks - a publishing format for reproducible computational workflows. In F. Loizides & B. Scmidt (Eds.), *Positioning and power in academic publishing: Players, agents and agendas* (pp. 87–90). IOS Press. https://eprints.soton.ac.uk/403913/

Kowalski, A. F. (2024). Stellar flares. *Living Reviews in Solar Physics*. https://arxiv.org/abs/2402.07885

NASA. (2014, August). Swift's ultraviolet/optical telescope (uvot) [Accessed: 2024-07-30].

Paudel, R. R., Barclay, T., Youngblood, A., Quintana, E. V., Schlieder, J. E., Vega, L. D., Gilbert, E. A., Osten, R. A., Peacock, S., Tristan, I. I., Feliz, D. L., Boyd, P. T., Davenport, J. R. A., Huber, D., Kowalski, A. F., Monsue, T. A., & Silverstein, M. L. (2024). A multiwavelength survey of nearby m dwarfs: Optical and near-ultraviolet flares and activity with contemporaneous tess, kepler/k2, swift, and hst observations. *Astrophysical Journal*. https://doi.org/https://doi.org/10.48550/arXiv.2404.12310

Ranjan, S., Wordsworth, R., & Sasselov, D. D. (2017). The surface uv environment on planets orbiting m dwarfs: Implications for prebiotic chemistry and the need for experimental follow-up. *The Astrophysical Journal*, *843*(2), 110. https://doi.org/10.3847/1538-4357/aa773e

Shields, A. L., Ballard, S., & Johnson, J. A. (2016). The habitability of planets orbiting M-dwarf stars., *663*, 1. https://doi.org/10.1016/j.physrep.2016.10.003

Shulyak, D., Reiners, A., Nagel, E., Tal-Or, L., Caballero, J. A., Zechmeister, M., Béjar, V. J. S., Cortés-Contreras, M., Martin, E. L., Kaminski, A., Ribas, I., Quirrenbach, A., Amado, P. J., Anglada-Escudé, G., Bauer, F. F., Dreizler, S., Guenther, E. W., Henning, T., Jeffers, S. V., . . . Pedraz, S. (2019). Magnetic fields in m dwarfs from the carmenes survey. *AA*, *626*, A86. https://doi.org/10.1051/0004-6361/201935315

Stetson, P. B. (1987). DAOPHOT: A Computer Program for Crowded-Field Stellar Photometry., *99*, 191. https://doi.org/10.1086/131977

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., . . . SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2