MURAJ
Minnesota Undergraduate Research & Academic Journal
z.umn.edu/MURAJ

Spatial and Temporal Identification of Extreme Ultraviolet Flares in Solar Active Regions

By Theo Mahaffey[1]

[1]*School of Physics and Astronomy, University of Minnesota - Twin Cities*

Abstract: An automatic method developed to perform flare-finding on Level 1 extreme ultraviolet (EUV) data from the Atmospheric Imaging Assembly (AIA) is described in detail. The data are spatially binned, preprocessed, and compared against a Gaussian white noise background. The methods of Berghmans and Clette (1998) provide the basis for the automated detection algorithm, and pertinent adjustments are made to address the high spatial resolution provided by AIA. Results for an observation of NOAA active region 12712 from 16:00-16:45 UTC on 2018/05/29 are presented. Future work may include investigation of the link between EUV active regions and periodicities found in various observables, including radio emissions and magnetic field. Accurate determination of the flare-frequency distribution for small flares is also a promising application of this method.

1. Introduction

Solar flares are of interest to those who study the sun and its impact on the Earth. These energetic releases of charged particles and electromagnetic radiation are closely connected with many other solar phenomena. For example, recurring jets observed in extreme ultraviolet (EUV) wavelengths have been linked to the magnetic field dynamics at the footpoints of these events (Joshi et al., 2017). Large flares are the drivers of a large part of what is collectively known as "space weather", so understanding the physics of these events is vital to predicting the ever-changing space weather conditions. In 1995, it was estimated by the United States NOAA that tens of millions of dollars are lost per year due to space weather. As science and society alike become more dependent on satellite technology for communication, discovery, and defense, our vulnerability increases as well. For this reason, many solar physicists are working to develop more accurate models to describe and forecast space weather (Baker, 1998).

Smaller flares are also crucial to our understanding of the processes that drive solar activity. The specific mechanism or mechanisms that heat the solar corona to several million Kelvin remain undetermined. The coronal heating problem has many proposed solutions, one of which involves the subset of very small flares, known as nanoflares (Parker, 1988). Nanoflares are considered to have comparable physical parameters and to follow the same process as flares and microflares, just with a lower total

energy (Hudson, 1991). If these small events occur with a sufficiently high frequency, it is possible that they provide enough energy to sustain the temperatures observed in the corona. Many studies have focused on determining the occurrence frequency of a certain subset of flares; some of their results are summarized in Figure 1.

However, identifying flares — especially small nanoflares — in vast datasets is a time-consuming and largely inefficient step of this research. An automated detection method has been applied to data from the Extreme-Ultraviolet Imaging Telescope (EIT), launched in 1995 (Berghmans & Clette, 1998). The higher-resolution images recorded by the newer Atmospheric Imaging Assembly (AIA) provide more valuable observations of smaller flares. This paper describes how the methods of Berghmans and Clette were adapted to develop a program that identifies and catalogs significant brightenings in high-resolution solar EUV data recorded by AIA, which is a tool that will aid a diverse field of future study, including investigations of additional characteristics of flaring regions and refinement of the flare frequency distribution.

2. Methods

2.1 General Characteristics of AIA Data

AIA, aboard the Solar Dynamics Observatory (SDO) records full-disk images of the sun in seven EUV wavelengths every 12 seconds (Lemen et al., 2012). This rapid sampling enables the observation of solar activity down to timescales on the order of seconds. The telescope has a 1.5-arcsecond spatial resolution but produces images with 0.6-arcsecond pixels, which enables analysis of events on very small length scales. Each pixel has a value, in arbitrary units, proportional to the intensity of solar radiation at a certain EUV wavelength integrated over the solid angle subtended by the pixel (Boerner et al., 2012).

2.2 The Automated Detection Algorithm

Berghmans and Clette (1998) developed a transient detection algorithm for data recorded by EIT. The framework of their algorithm can be applied to the identification of significant brightenings in AIA data, with modifications necessitated by the higher cadence and spatial resolution of AIA. EIT images consist of 2.6-arcsecond pixels providing a 5-arcsecond spatial resolution — more than three times coarser in both dimensions than the measurements performed by AIA (Delaboudinière et al., 1995). One of the features added to the original algorithm to compensate for the small-scale fluctuations resolvable by AIA was the use of macropixels.

2.2.1 Macropixels

"Macropixel" refers to one unit of a spatially binned image, specifically the average over a three-pixel by three-pixel region of the original AIA image. The use of macropixels reduces the statistical fluctuations caused by the high spatial resolution of the data. Figure 2 displays the scheme by which macropixels are created from AIA images. The result is a report of EUV emissions with a 1.8 arcsecond spatial resolution, which is still an improvement over the resolution of EIT images.

By creating macropixels from each of many subsequent images of the same region, an interval of AIA data is constructed. Each macropixel, M^{xy}, has a brightness time series, or lightcurve, that is represented as $\vec{b}^{xy} = \left[b_0^{xy}, b_1^{xy}, \ldots, b_{N-1}^{xy}\right]$. The timestep, determined by the instrument cadence, is 12 seconds, and the number of datapoints, N, is

directly controlled by the length of the interval being analyzed.

2.2.2 Preprocessing

The algorithm ultimately utilizes data obtained by preprocessing macropixel lightcurve values as outlined in Berghmans and Clette (1998). The first step in the preprocessing is to compute a boxcar average for each macropixel and subtract it from the original lightcurve. This step removes background trends that act with a time profile much longer than that of most flares. Letting w be the boxcar width in either direction, the resulting residual lightcurve is $\vec{r}^{xy}$ with

$$r_i^{xy} = b_i^{xy} - \frac{1}{2w+1}\sum_{j=i-w}^{i+w} b_j^{xy} \qquad (1)$$

The first and last w terms of $\vec{b}^{xy}$ are necessarily discarded in this step; to preserve consistency in the notation, $\vec{r}^{xy} = [r_w^{xy}, r_{w+1}^{xy}, \dots, r_{N-w-1}^{xy}]$. The last preprocessing step converts each residual lightcurve value to a z-score via the process outlined in Eqns. 2-4, which streamlines comparison to the statistical model discussed below in the Theory section. The resulting time series $\vec{z}^{xy}$, whose entries are defined in Eqn. 4 below, is referred to as the "signal".

$$\overline{r^{xy}} = \frac{1}{N-2w}\sum_{j=w}^{N-w-1} r_j^{xy} \qquad (2)$$

$$\sigma^{xy} = \sqrt{\frac{1}{N-2w}\sum_{j=w}^{N-w-1}\left(r_j^{xy} - \overline{r^{xy}}\right)^2} \qquad (3)$$

$$z_i^{xy} = \frac{r_i^{xy} - \overline{r^{xy}}}{\sigma^{xy}} \qquad (4)$$

2.2.3 Event Construction

The next step is to identify bright regions in each image using the calculated signal. To achieve this, Berghmans and Clette (1998) defined two thresholds: one for detection and one for connection. The detection threshold sets the minimum signal value for a macropixel to be flagged as bright and is denoted as q_d. When the signal value of a macropixel, M^{xy}, at time i is found to be above q_d, the surrounding region is investigated further to determine the full spatial extent of the brightening. This is done by comparing the adjacent macropixels against the connection threshold q_c. If the signal value of any neighboring macropixel exceeds q_c, its immediate neighbors are checked for connection as well. This continues recursively until the full spatial extent of the brightening has been covered. Due to the smaller-scale fluctuations present in AIA images, the following restriction was added to prevent false positives: if no macropixels can be connected to the original detection, it is discarded to limit the prevalence of false positives. Groups of two or more bright macropixels are combined into "clusters."

Once clusters have been created for each image in the interval, the program utilizes the instantaneous clusters to reconstruct "events" with a meaningful time profile. Events are the final product of the detection algorithm. As another measure to prevent false positives due to the higher cadence of AIA, a bright region (cluster) must be present in at least two consecutive images for that activity to be considered a significant event. To enact this requirement, the clusters found in one image are checked against the clusters in the following image. Any clusters across the two images that overlap spatially by at least one macropixel are joined into an event. Consecutive images are checked for the entire interval, resulting in the final list of events present in that interval. This collection of events fully describes the brightnesses,

times, and locations of the flares found in a given interval of AIA data.

3. Theory

3.1 Background Model

The calculations that assign meaning to the threshold values rely on a Gaussian white noise background model. The Level 1 AIA data used in this study are directly proportional to the energy output by the sun and received by the instrument. The filters and optics of the telescope were optimized to have a very narrow bandpass around the desired wavelength for each channel (Boerner et al., 2012). Therefore, the recorded data are, to a good approximation, directly proportional to the number of photons received in the short (< 3 seconds) exposure times of each image (Lemen et al., 2012). This allows for Poisson statistics to be applied. For a large number of photon counts, which is expected from an active region producing flares, the Poisson model is approximated very well by a normal distribution. As a result, it is expected that the background behavior of each macropixel's signal can be described by a Gaussian distribution.

To test the validity of this assumption, histograms of the residual lightcurve and signal values for several arbitrarily chosen macropixels were generated and compared to a Gaussian distribution. Macropixels were selected from regions with a range of brightness levels to ensure that no biases result from the application of this model. Figure 3 presents such histograms for two macropixels. The plots on the left show the count distribution of r^{xy}, and the normalized histograms on the right plot the probability distribution of z^{xy} for each macropixel. If the signal values that result from preprocessing can be interpreted as a valid representation of z-scores for a variable that follows a normal distribution, their probability distribution should closely resemble the unit Gaussian plotted alongside them. Figure 3 demonstrates that this is the case.

3.2 Probability of False Positives

Having confirmed the Gaussian nature of the background activity of the AIA data, it is now possible to estimate the probability of false detection and establish meaningful thresholds for significance. Given a threshold q, the probability of a background fluctuation producing a signal that exceeds the threshold is given by

$$P(z \geq q) = 1 - \int_{-\infty}^{q} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz \qquad (5)$$

When $q = q_d$, this quantity is the likelihood of a single macropixel producing a false positive detection in a single image. By requiring that at least one macropixel be connected to a detected macropixel, the probability of a false detection being reported as a cluster (C_F) is reduced to

$$P(C_F) = P(z \geq q_d) \cdot P(z \geq q_c) \qquad (6)$$

The probabilities can be multiplied in this way because of the Gaussian white noise model to which the signal is being compared. There is no autocorrelation in such a model, meaning that the activity of each macropixel is assumed to be independent of the activity of all other macropixels. This applies to spatially separated macropixels in the same image as well as the same macropixel temporally separated across images. Thus a further step taken to reduce false positives is to require that a cluster appears in at least two consecutive images. This reduces the probability of a falsely detected event (E_F) to

$$P(E_F) = P(C_F)^2 \qquad (7)$$

Results are given in the next section for an interval that was tested with $q_d = 6$ and $q_c = 4$. These values give a false detection probability of $P(E_F) = 9.76 \times 10^{-28}$, which is quite small. This is necessary to counter the vastness of the datasets produced by AIA. In the small region and short time used as a test interval, there are 1.1×10^7 opportunities for a false detection. This number quickly increases for larger regions and longer intervals.

4. Results

The program produces a database of the times, locations, and brightnesses of all flares discovered in the signal. This database can be further applied to investigate, for example, the thermal properties of flares or the correlation between flare size and occurrence frequency. The program is also capable of generating movies of the active region with the significant events highlighted. An example frame is given in Figure 4. To test the program, an interval and region studied previously (Duncan et al., 2021) was chosen: 2018-05-29 16:00-16:45 UTC. Two wavelengths, 171 angstrom (Å) and 335 Å data from NOAA active region 12712 were investigated with various detection and connection thresholds (NOAA, 2008). The largest X-ray flare studied by Duncan et al. (2021) occurred around 16:20. At 16:20, the flare finding program with $q_d = 6$ and $q_c = 4$ identifies three simultaneous brightenings in the 171 Å AIA data. At no other time in the interval are three brightenings identified simultaneously.

The values of q_d and q_c used for analysis are dependent upon the study goals. If one is investigating only the brightest flares, a q_d of 8 or 9 could be appropriate, whereas a q_d of 4 or 5 would be more suited for a nanoflare study. The user has control over the definition of a flare and the ability to tune it to suit any specific analysis. For general purposes, $q_d = 6$ and $q_c = 4$ were found to provide a reasonable level of detection comparable to identifying flares by eye. With these thresholds, 71 events were identified in the 45-minute interval.

5. Conclusion

This automated flare detection program is a significant improvement on an otherwise cumbersome step of heliophysics research. Its automated nature and very small probability of false positives enable many more events and phenomena to be reliably studied in detail. In the absence of such a program, time is wasted on the imprecise task of identifying flares by eye. Being able to quantitatively determine the spatial and temporal extent of events of interest allows for more accurate links to be made between various observed quantities and the processes that produce them. A reliable method of flare finding could also be employed to better understand the frequency with which flares occur, specifically how the size and time scales of flares relate to the frequency of their occurrence. A conclusive result on this matter would shed light on the proposition of nanoflares as a solution to the coronal heating problem (Parker, 1988). Our algorithm may provide the level of detection necessary to support the theory.

6. Acknowledgements

The author is grateful for the support and guidance of Professors Cynthia Cattell and Lindsay Glesener. Additionally, Reed Masek provided many insights during helpful discussions. This work was supported by an Undergraduate Research Scholarship awarded by the University of Minnesota Office of Undergraduate Research.

Figures

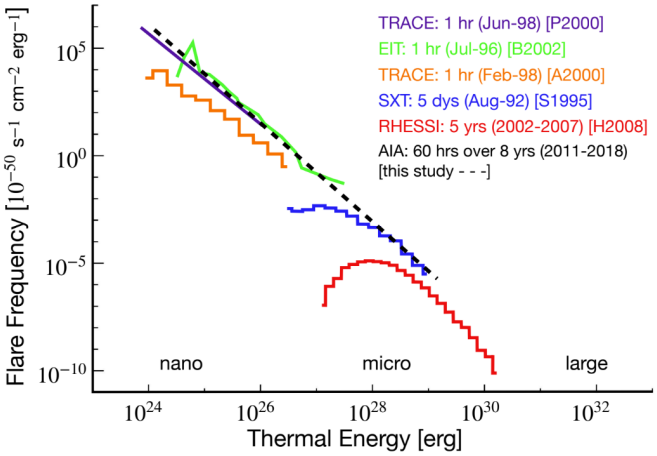

Figure 1. Compilation of solar flare energy distributions determined by various missions. Figure taken from Purkhart and Veronig (2022).

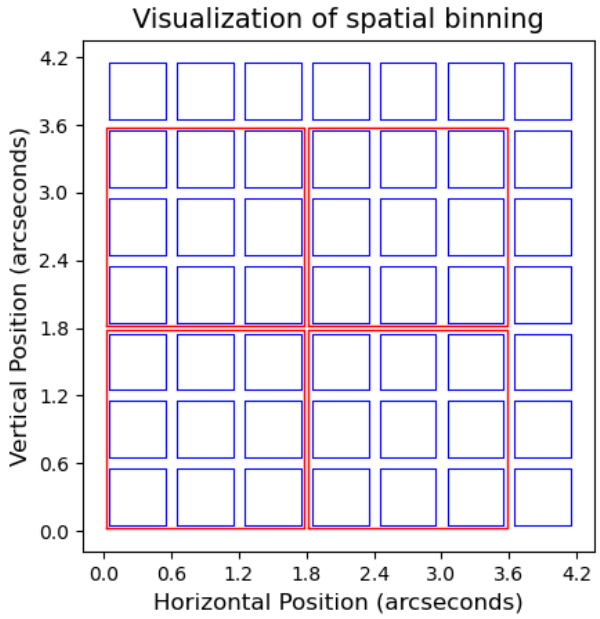

Figure 2. A conceptualization of the arrangement in which macropixels (red) are created by averaging squares of nine individual pixels (blue). Macropixels do not overlap, so the number of rows and columns of pixels in a spatially binned image must be a multiple of three. One or two rows/columns may be discarded, but the effect is insignificant.

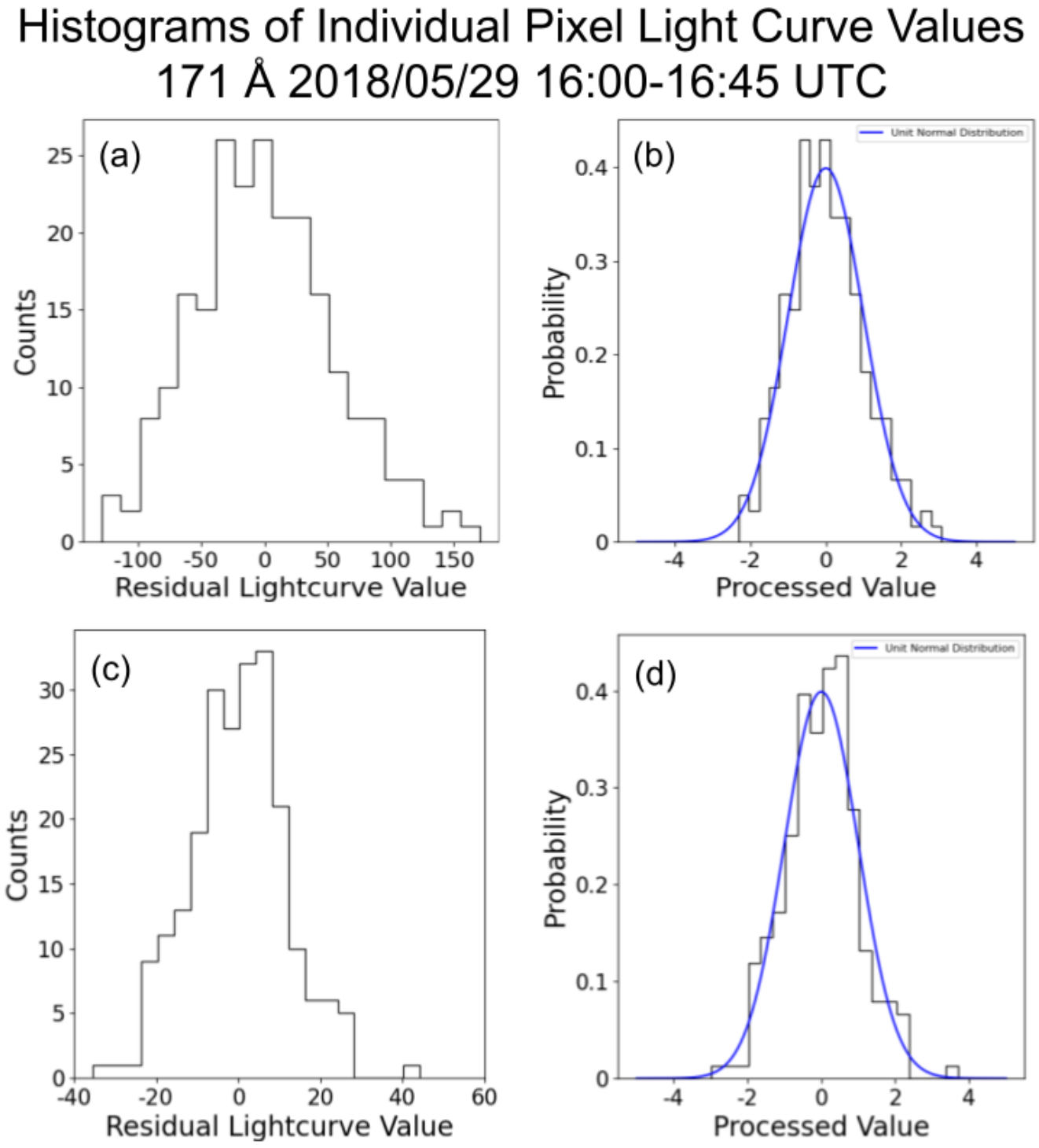

Figure 3. Histograms of the residual lightcurve value, r^{xy}, and the processed value, z^{xy}, for selected macropixels in the test interval. (a) and (b): A macropixel in a region of medium brightness. (c) and (d): A macropixel in a highly bright region. The Gaussian nature of the histograms supports our interpretation of the processed values as z-scores.

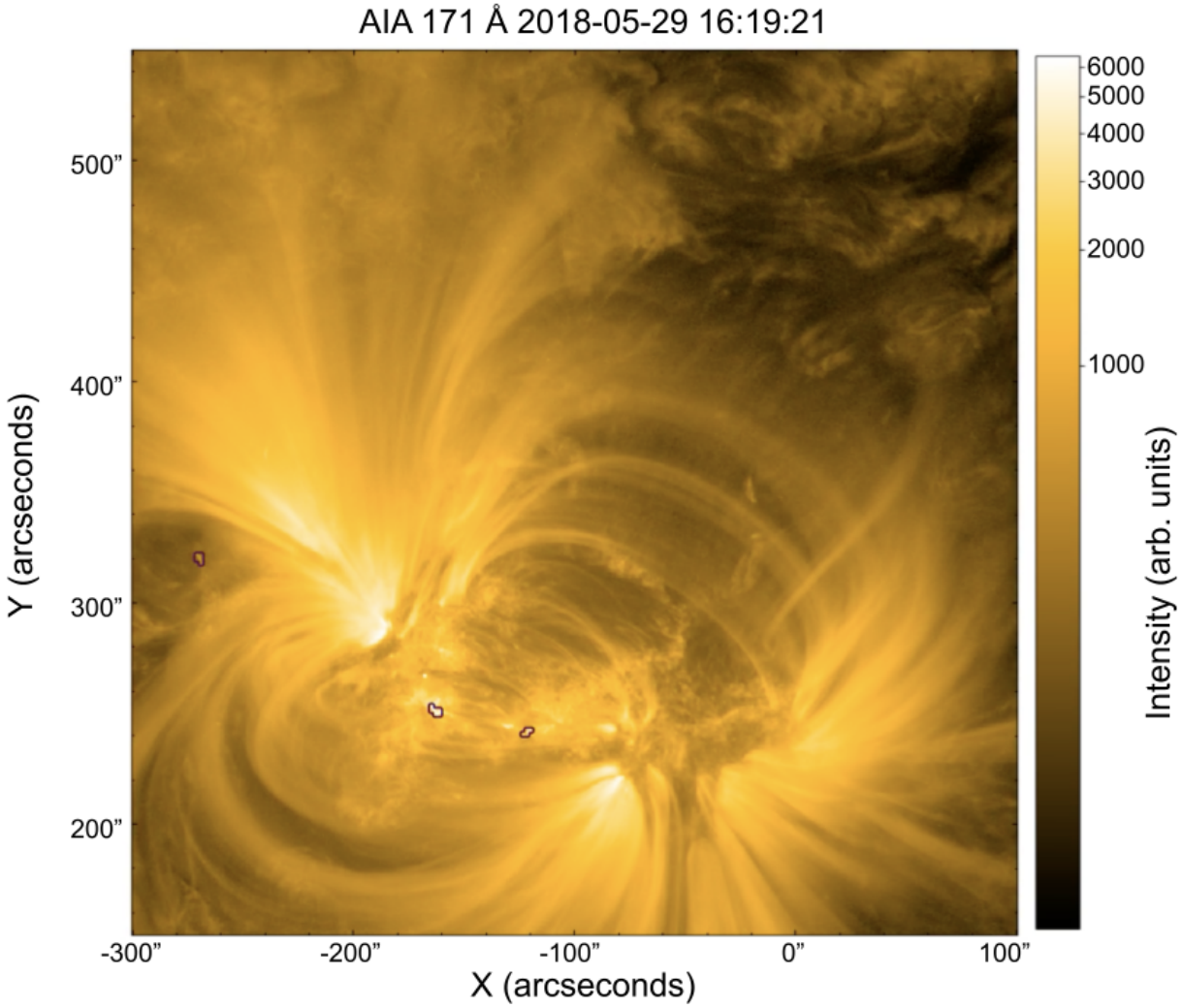

Figure 4. An example frame of 171 Å data from the test interval with three identified events outlined in purple (lower left). The detection and connection thresholds employed were 6 and 4, respectively.

References

Baker, D.N. (1998). What is space weather? Advances in Space Research, 22(1), 7-16.

Berghmans, D., Clette, F. (1999). Active region EUV transient brightenings - First results by EIT of SOHO JOP 80. Solar Physics, 186(1), 207-229. DOI: 10.1023/A:1005189508371

Boerner, P., Edwards, C., Lemen, J., Rausch, A., Schrijver, C., Shine, R., Shing, L., Stern, R., Tarbell, T., Title, A., et. al. (2012). Initial calibration of the Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO). Solar Physics, 275, 41-66. DOI: 10.1007/s11207-011-9804-8

Delaboudinière, J. P., Artzner, G. E., Brunaud, J., Gabriel, A. H., Hochedez, J. F., Millier, F., Song, X. Y., Au, B., Dere, K. P., Howard, R. A., et. al. (1995). EIT: Extreme-ultraviolet imaging telescope for the SOHO mission. Solar Physics, 162, 291-312. DOI: 10.1007/BF00733432

Duncan, J., Glesener, L., Grefenstette, B., Vievering, J., Hannah, I., Smith, D., Krucker, S., White, S., Hudson, H. (2021). NuSTAR observation of energy release in 11 solar microflares. The Astrophysical Journal, 908(29), 1-16. DOI: 10.3847/1538-4357/abca3d

Hudson, H. S. (1991). Solar flares, microflares, nanoflares, and coronal heating. Solar Physics, 133, 357-369. DOI: 10.1007/BF00149894

Joshi, N., Chandra, R., Guo, Y., Magara, T., Zhelyazkov, I., Moon, Y., Uddin, W. (2017). Investigation of recurrent EUV jets from highly dynamic magnetic field region. Astrophysics and Space Science, 362(10), 1-14. DOI: 10.1007/s10509-016-2983-x

Lemen, J., Title, R., Akin, D., Boerner, P., Chou, C., Drake, J., Duncan, D., Edwards, C., Friedlaender, F., Heyman, G., et. al. (2012). The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO). Solar Physics, 275, 17-40. DOI: 10.1007/s11207-011-9776-8

NOAA, US Dept of Commerce, Space Weather Prediction Center, USAF. (2008). Solar Region Summary. http://swpc.noaa.gov/products/solar-region-summary

Parker, E.N. (1988). Nanoflares and the solar X-ray corona. Astrophysical Journal, 330, 474. DOI: 10.1086/166485

Purkhart, S., Veronig, A. (2022). Nanoflare distributions over solar cycle 24 based on SDO/AIA differential emission measure observations. Astronomy and Astrophysics, 661, A149. DOI: 10.48550/arXiv.2203.11625