Analysis of Solar Flare Events

COSC 3337: Data Science I

Ly Ha
Computer Science
University of Houston
Houston, TX USA
lythituyetha@gmail.com

Erica Hay
Computer Science
University of Houston
Houston, TX USA
Ericakimhay03@gmail.com

Khuong Ngo
Computer Science
University of Houston
Houston, TX USA
khuongtngo5@gmail.com

ABSTRACT

Our team, supplied with datasets from the Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI), embarked on the creation and implementation of a system named Helios. The primary aim of Helios is to efficiently process spatial-temporal data, perform mapping, identify hotspots, and conduct change analysis using information derived from RHESSI on high-intensity solar flare events. To achieve this objective, a thorough comprehension of the data and its associated variables is crucial. The data given to us was split into 2 datasets. The first describes solar flare events that take place between the years 2004-2005 while the second describes events between the years 2015-2016. Additionally, each of these datasets is characterized by attributes such as the date of the flare occurrence, flare start time, flare peak time, flare end time, duration of the flare, peak count rate, total counts, energy band, x position in arcseconds, y position in arcseconds, and radial distance in arcseconds. Comparing how these variables fluctuate during two distinct periods allows us to deepen our understanding of why solar flares occur.

CCS CONCEPTS

• Mathematics of computing • Probability and Statistics •Multivariate statistics

KEYWORDS

Solar Flares, Hotspot Estimation, Change Analysis, NASA, Intensity Estimation, Spatiotemporal Analysis, Temporal Analysis

ACM Reference format:

Ly Ha, Erica Hay and Khuong Ngo. 2023. Analysis of Solar Flare Events: COSC 3337: Data Science I. In *Proceedings of Data Science I at the University of Houston., Houston, TX, USA, 8 pages.*

1 Analysis of Temporal Data and How it Affects Frequency and Intensity of Solar Flares

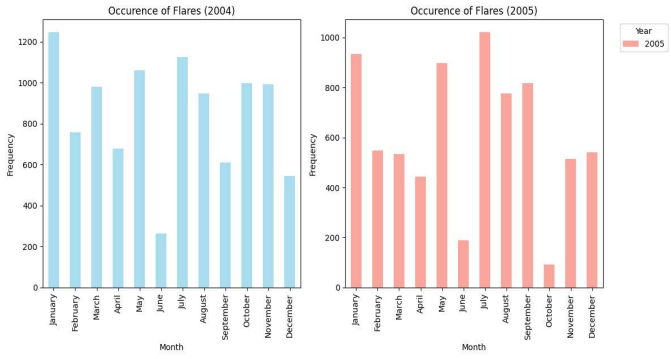


Figure 1: Occurrences of Flares based on a Monthly Distribution for 2004 and 2005

We begin our analysis by examining the dates of flare occurrences. By comparing the frequency of flare occurrences between 2004-2005 and 2015-2016, we aim to identify trends to understand the frequency and distribution of solar flares. In **Figure 1**, we can observe that the 2004 and 2005 plots exhibit similarity in the monthly distribution of solar flares. Both plots reveal peaks in January and July, followed by dips in June and December. This consistent pattern may suggest a strong influence of the Sun's magnetic field, which undergoes regular variations throughout the solar cycle. As stated by NASA, "The magnetic field lines near sunspots often tangle, cross, and reorganize. This can cause a sudden explosion of energy called a solar flare". However, it is worth mentioning that the frequency of solar flares in October 2005 was abnormally low compared to other months in the two years. This can be characterized as an outlier in our data. The 2015 and 2016 plots reveal a significant difference in frequencies of solar flares. As shown in **Figure 1**, the largest amount of flare occurrences for 2015 is near 1400 while the largest for 2016 is only around 500. Moreover, there appears to be no flares recorded for March and April in the year of 2016. This drastic change can be explained by concepts known as the solar maximum and minimum of a solar cycle. As described by the University Corporation for Atmospheric Research, the solar cycle is where "the Sun resets with a new, smooth magnetic field, that field is reversed from the way it had just been. The Sun's North and South Magnetic Poles trade places. So it really takes about 22 years (two 11-year cycles) for the Sun's magnetic field to go through a complete cycle and end up back where it started". The first 11-year cycle is the height of the Sun's activity cycle, known as the solar maximum, and is a time of greatly increased solar storm activity. The last 11-year cycle is the opposite, known as the solar minimum, and is a time of decreased solar activity. By examining the four plots, it becomes clear that the frequency of solar flares was significantly higher in the years 2004, 2005, and 2015 compared to the plot for 2016. This observation aligns with the fact that the solar maximum occurred between 2004 and 2015, a period of 11 years. The absence of solar flares in 2016 can be attributed to the onset of the solar minimum. This pattern further reinforces the influence of the Sun's magnetic field and solar cycle on the monthly distribution of solar flares.

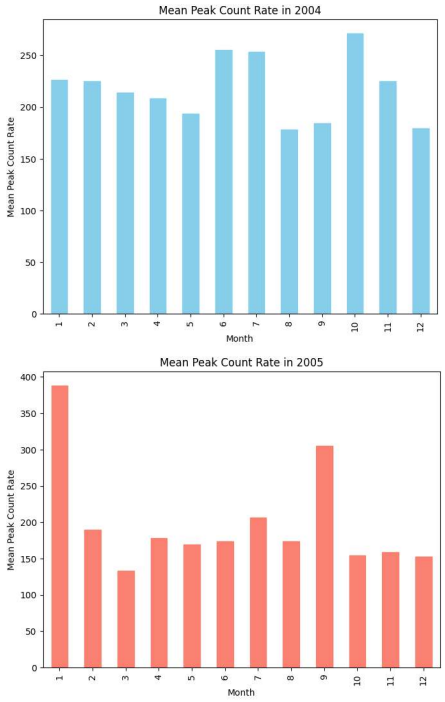


Figure 2.1 and 2.2: Distribution of Mean Peak Count Rates in 2004 & 2005: Plot 1 shows two prominent peaks in 2004 in the last half of the year, contrasting Plot 2's peaks in the first and last quarter of 2005.

We continue our analysis by looking at the distribution of the mean peak count rates for both datasets. As shown in **Figure 2**, there are two peaks where the mean peak count rate was the highest, while the rest of the data has a relatively equal distribution for the 2004 data. The data has a starting mean peak count rate of 226.32 for January and slowly decreases to a mean peak count rate of 193.62 in May. The first peak of the data occurs in June and July, where the mean peak count rates are 255.34 and 253.29, respectively. The months of August and September experienced a steep drop in the mean peak count rates, 178.40 and 284.43, and then another peak occurred in October, 271.27. After October, the mean peak count rates steadily decreased for the remainder of the year. In contrast, the 2005 data displays a distinct distribution. In the second plot of Figure 3, there are two prominent peaks, January, 388.05, and September, 300. The rest of the means are evenly distributed around 150-170. While both years exhibit a relatively even distribution of mean values, the overall mean values in 2005 are much lower than the overall mean values for 2004.

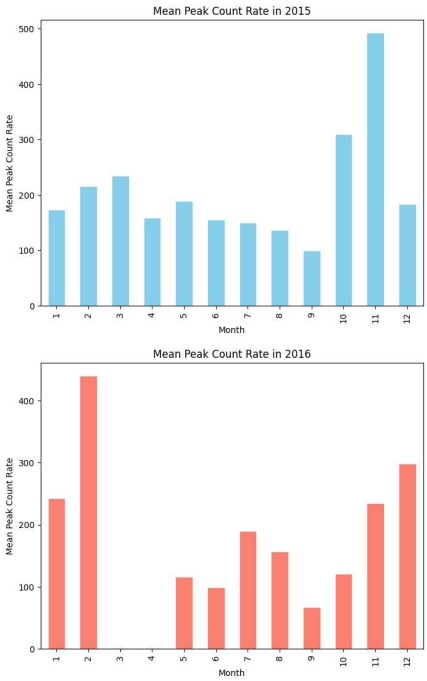


Figure 3.1 and 3.2: Contrasting Mean Peak Count Rates Distributions in 2015 & 2016: Plot 1 shows fluctuating mean values throughout the year of 2015, while plot 2 shows distinct peaks in the first quarter of the year of 2016, contributing to a less evenly distributed pattern.

The 2015-2016 datasets show a large difference in mean peak count rate distributions for both years. In **Figure 3**, the mean values started on the lower end for the 2015 data. January has a mean value of 171.72 and the data experiences a slight increase until March, where the mean value is 232.81, and then a steady decrease until September, 97.89. In October and November, the data saw a very large increase in mean values, 307.75 and 491.24, and then another steep decrease in December, 182.14. In 2016, the data experienced a very different distribution in the mean peak count rates, as shown in the second plot of **Figure 3**. January has a mean value of 241. 64, and then February experienced a sudden increase, where the mean was 439.06. The rest of the mean values are fairly lower and saw two smaller peaks, one in July and the last in December. Compared to the 2015 data, the mean values of 2016 are less evenly distributed and experience more peak values. However, the mean values are also slightly higher than the ones seen in 2014.

The two datasets exhibit distinct distributions of their mean values. While the 2004-2005 dataset shows evenly distributed data, the 2015-2016 dataset presents variations in mean values with notable peaks and troughs. The overall mean peak count rates for 2004-2005 are much higher than the second data set. As shown in both plots of **Figure 2**, most of the mean values are concentrated around 200, while both plots in **Figure 3** mainly exhibit values below 200. The disparities may be influenced by

factors such as fluctuations in solar activity, which impact the occurrence and intensity of solar flares.

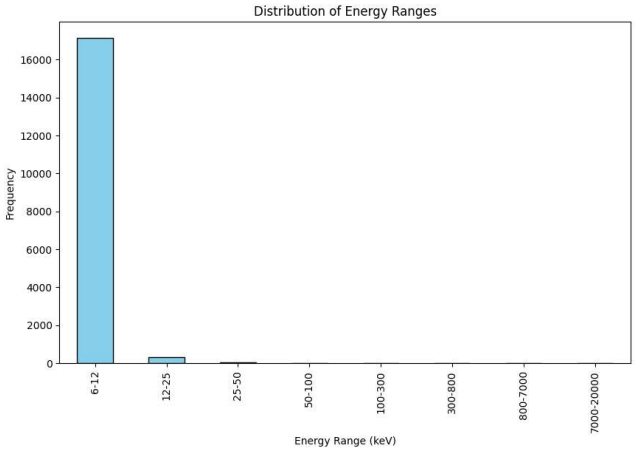
1.1 Differentiation on Averages and Variability

As we delve deeper into the attributes of solar flares, it is essential to analyze thoroughly the quantitative aspects captured by the two datasets. By examining key statistical measures for various variables, we can enhance our comprehension of the temporal, spatial, and energetic dimensions of these solar events. The patterns observed in flare occurrences, which aid our research of implicit factors shaping these events, may reinforce the previously mentioned notion of the 22-year solar cycle.

We now shift our focus to the numerical insights provided by the temporal variables of our datasets. These time-based attributes include the date and duration of solar flare events, where the date attribute is split into the year, months, and days. We can begin by computing basic statistics such as the mean and standard deviation to reveal the underlying distribution of values. The first dataset exhibits a mean date of November 6, 2004, with a corresponding standard deviation of 207 days. In contrast, the second dataset shows a mean date of July 14, 2015, with a standard deviation of 178 days. The second data set has a notably lower mean 'month' value, indicating a concentration of observations in the earlier months of the year and a distribution that is skewed right. As previously stated, this is caused by the solar cycle and how magnetic poles are flipped. Additionally, the standard deviation of the second dataset is lower than the first, indicating that the dates are less varied from the mean date. This tells us that the timing of flare occurrences is less consistent in 2015-2016 than it was in 2004-2005. In considering the duration variable, Dataset 2 has a slightly higher mean duration (509.91 seconds) compared to Dataset 1 (496.87 seconds). However, Dataset 2 exhibits a lower standard deviation (432.23 seconds) in contrast to the higher variability seen in Dataset 1 (444.05 seconds). This suggests that while the average solar flare duration is slightly longer in Dataset 2, the durations in Dataset 1 are more spread out. We can infer from this finding that, on average, high-intensity flare events are shorter in duration.

Moving on to spatial variables, 'x position' and 'y position' show considerable differences. It should be noted that these variables are measured in arcseconds from the center of the sun, where the center is denoted as (0,0). Dataset 1 has negative mean values for both variables (-22.70 and -32.64, respectively), indicating a tendency towards the left and downward directions. In contrast, Dataset 2 has positive mean values (8.90 and 12.72, respectively), suggesting a shift towards the right and upward directions. This implies that the active region, the area marked by the presence of strong magnetic fields, is generally positioned near the left side of the sun's equator for 2004-2005. Alternatively, the general position of the active region for 2015-2016 is to the right side of the Sun's equator. The standard deviations for these variables are higher in Dataset 1, indicating greater dispersion in spatial positions.

As we conclude our examination of the temporal and spatial dimensions, we now turn our focus to the energetic variables. The peak count rates variable provides insight into the rate at which photons are most observed in both datasets. The higher mean peak count rate in Dataset 1 (217.46 counts per second) compared to Dataset 2 (193.24) suggests that, on average, solar flares in Dataset 1 exhibit more intense bursts of activity. However, it is crucial to note that the substantially higher standard deviation in Dataset 1 (851.79 counts per second) indicates a wider variability in peak count rates compared to Dataset 2 (547.45 counts per second). This increased variability implies a broader range of peak count rates, with some solar flares in this dataset exhibiting exceptionally high levels of activity. The reason for this could be influenced by, once again, the magnetic field strengths. The total count of events within the specified energy range provides an overall measure of the intensity of solar flare events in both datasets. Dataset 1 demonstrates a higher mean total count (4.41e+05) compared to Dataset 2 (2.78e+05), indicating, on average, a greater cumulative intensity of solar flares in Dataset 1. The notably higher standard deviation in Dataset 1 (3.37e+06) compared to Dataset 2 (1.29e+06) suggests a wider spread of total counts in Dataset 1. This greater variability in total counts in Dataset 1 may be attributed to a more diverse range of solar flare intensities or a higher frequency of extreme events within this dataset. The energy band range variable characterizes the uppermost energy range in which the flare is observed, measured in kilo electron volts. The spectrum spans from the lowest energy band range of 6-12 to the highest at 7000-20000. Despite the striking similarity in the mean and standard deviation as seen in **Figure 4**, Dataset 1 exhibits notably higher frequency compared to Dataset 2. This discrepancy may be attributed to the frequent occurrence of low-energy flares, contrasting with the infrequent nature of high-energy flares in both datasets.



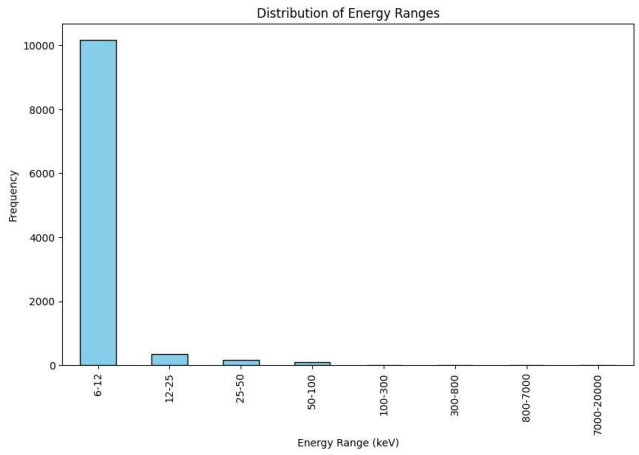


Figure 4.1 and 4.2: Frequencies of which the highest energy band of a flare is observed in 2004-05 and 2015-16 (respectively)

While the computations of mean and standard deviation offer valuable insights into the average characteristics, variability, range, and distribution of solar flares, a comprehensive understanding of their behavior requires further investigation. This involves the simultaneous interpretation of multiple variables and a detailed analysis of the trends they generate. Having already explored the interplay between temporal data and energy variables, our next step involves delving into the correlation between spatial variables and energy characteristics.

2 Comparing Intensity through Spatial Variation

In order to evaluate the intensity of the solar flare events, our team developed two methodologies that calculate intensity based on different factors: the first, relying on the total number of photons detected during a flare, and the second, utilizing the energy band's average and the flare's duration. Our analysis also involves the calculation of intensity values using seven query points. Through the visualization of outcomes using four intensity maps, we analyze the occurrence of solar flares during the initial four months (months 1, 2, 3, and 4) and the concluding four months (months 21, 22, 23, and 24) across both datasets. The two different intensity scores allow us to observe unique insights into spatial distribution and total intensity of the solar flares on the Sun's surface.

2.1 Methodology for Intensity Calculations

Before delving into the analysis, let's delve deeper into the methodology behind calculating intensity scores in the first and second methods. To kick things off, our team strategically identified seven key query points by mapping the Sun's surface onto a grid. These points were selected to ensure an even distribution, including three along the equator (center, negative x-axis, and positive x-axis) and two in both positive and negative regions along the y-axis, resulting in regions four, five, six, and seven. Each region was assigned specific ranges, measuring 402 arcseconds in length and 670 arcseconds in width. We can observe these regions in **Figure 5**.

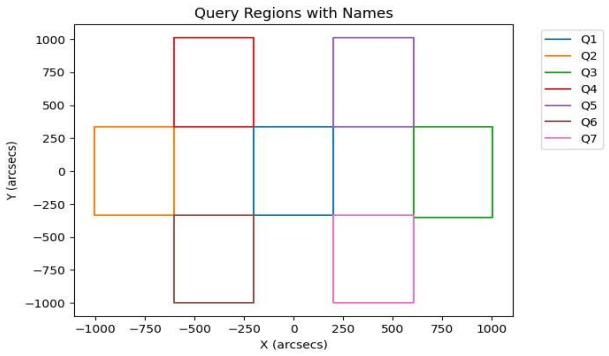


Figure 5: Query Regions with names, as specified above.

In the realm of Method 1, we gauged the intensity of solar flare events in distinct sun regions using kernel density estimation with the Gaussian kernel function written in **Figure 6**.

$$K(x, y, X, Y, \sigma) = exp(-\frac{(x-X)^2 + (y-Y)^2}{2\sigma^2})$$

Figure 6: Gaussian Kernel Function

This approach involved calculating the contribution of each solar flare event to intensity at a given query point, considering spatial proximity. A dedicated function is iterated through the events, accumulating contributions to intensity. The ensuing intensity distribution was visualized through kernel density estimation on spatial coordinates, resulting in log-densities that were later exponentiated to create a density map.

Shifting gears to Method 2, we followed a similar trajectory, employing kernel density estimation with a Gaussian kernel function. However, this method introduced a nuanced intensity calculation that factored in the duration and energy of each solar flare event, as shown in **Figure7**. Notably, the average energy within the specified energy band played a crucial role. mirroring Method 1, kernel density estimation was used to exponentiate log densities, yielding actual densities for visual representation.

$$Intensity = \sum (duration_i * energy_{avg,i} * GaussianKernel(X_q, Y_q, X_i, Y_i, \sigma))$$

Figure 7: Intensity Equation for Method 2

2.2 Analysis of Spatial Variability on Density/Intensity Estimation

In the analysis of the 2004-2005 dataset, the initial method divulged a persistent trend of heightened intensities across all query regions during the first and last quartiles of the temporal span. Notably, query region two emerged as the focal point of the

highest total count intensity with a value of 354,646,021.83 for the first four months. Regions four and seven also display relatively high intensities, 46,214,972.71 and 18,958,793.58, respectively, suggesting localized concentrations of significant solar flare events. Conversely, the intensity values of the remaining regions are less noteworthy, but regions five and six exhibit moderate intensities while regions one and three returned the lowest values among the seven, suggesting that there is a disparate distribution. The final four months show distinctive observations, where region two maintains the highest intensity, 278,024,534.21. Despite this, the intensity values experienced a considerable decrease compared to the initial four months. Noteworthy, nearly all query regions experienced a large decline in intensity values, with an exception region six which experienced an intensity value of 59,623,837.94, ranking it as one of the highest values of the seven.

The application of the second method on the same dataset produced similar results, where the intensity values for each query region mirror the results obtained using method one. Query region two displayed the highest intensity value, 6,403,524.02 during the initial four months. However, regions one and six showed relatively higher intensities, in contrast to method one's results for regions four and seven. The remaining regions showed lower intensity values. The last four months of the data show a similar pattern to method one, where the overall intensity values decreased across multiple query regions. The consistent decline in intensities using both methods implies an overall decrease in solar flare activity between the first four months and the last four months.

In Contrast to the 2004-2005 dataset, the 2015-2016 dataset yielded dissimilar patterns, particularly in the results using the first method. For the first four months, query region three experienced the highest intensity value, 292,559,164.29, deviating from the first dataset. Query regions two and five displayed relatively high intensities, while regions four and seven had much lower values compared to the 2004-2005 dataset. The results generated using method two are similar, with region three having the highest intensity, 9,331,348.80, but regions one and six show relatively high intensities. The last four months of the dataset exhibit similar observations for both methods, with all regions experiencing a decrease in intensity.

When comparing intensity values between the two datasets, similarities emerged, particularly in the significant decrease observed during the last four months across all query regions. Notably, the 2004-2005 dataset showcased the highest intensity in region two for both methods, while the 2015-2016 dataset exhibited the highest intensity for both methods in region three. However, a substantial difference is evident in the magnitude of the intensity values, with the 2015-2016 dataset displaying lower values for both methods.

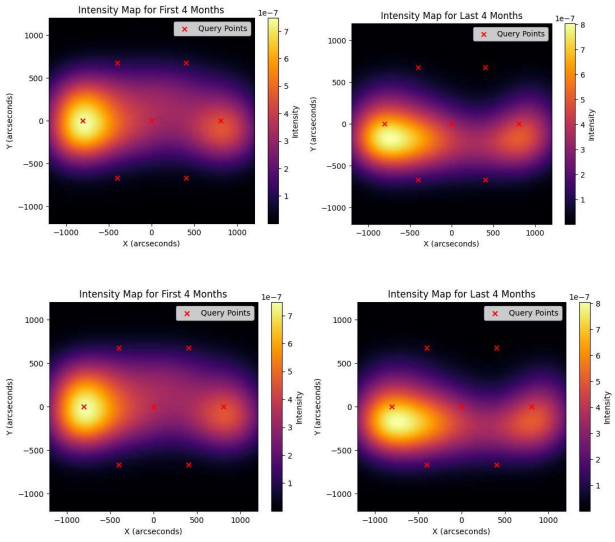


Figure 8.1-8.4: Density/Intensity Maps for 2004-05 Dataset in order from left to right. Map 1 – Method 1, Map 2 - Method 2. Map 1 – Method 2, Map 2 – Method 2.

In addition to computing individual intensity values for each of the seven query regions, we generated four density maps to provide a visual representation of the results for each dataset. Map one corresponds to the initial four months, while map two corresponds to the last four months (**Figure 8**). Each methodology is associated with a set of corresponding maps, offering a comprehensive visualization of the solar flare intensity distribution across the timeframes.

In the 2004-2005 dataset, both methodologies, namely the total number of photons and the energy band's average with flare duration, consistently revealed high-intensity regions within specified x and y ranges. As shown in **Figure 8.**1, the concentration of the highest intensities in query region two, as corroborated by intensity values, underscores the reliability of these findings. The spatial analysis brought to light a noteworthy prevalence of high-intensity areas within the x range of [-1000, 1000] arcseconds and the y range of [-500, 500] arcseconds. The congruence in patterns observed with both methods enhances the robustness of the identified regions of interest.

A nuanced distinction emerged in the y-axis, where the first method exhibited heightened intensity around the y range of [0, 500], while the second method demonstrated higher intensity around the y range of [0, -500], as shown in **Figure 8.3**. These nuanced variations highlight those intensities, whether measured by a high count of photons or the average of the energy band with flare duration, predominantly cluster around the solar equator. Utilizing visualizations, it becomes evident that the total intensities for both methods are comparable, with peak intensities for maps one and two reaching 7e-7 and 8e-7, respectively.

These subtle disparities underscore the potential influence of methodological choices on the precise localization of intense solar flare events. It is imperative to consider these variations when interpreting and selecting methodologies for solar activity

analyses, particularly regarding the accurate pinpointing of high-intensity regions.

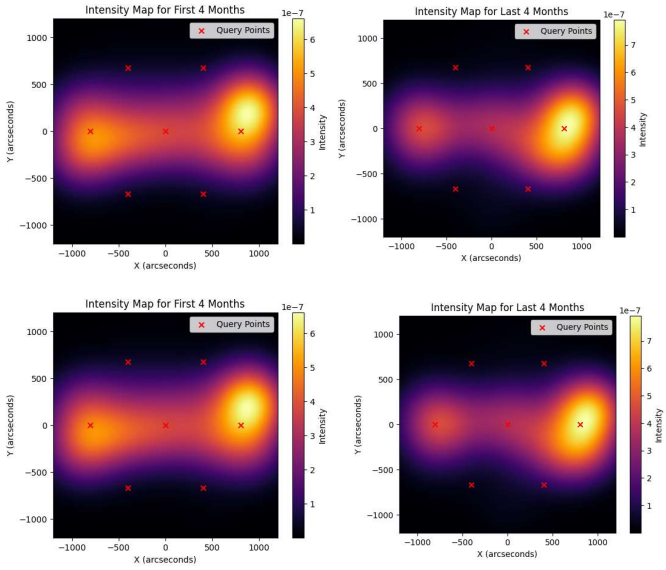


Figure 9.1-9.4: Density/Intensity Maps for 2015-16 Dataset in order from left to right. Map 1 – Method 1, Map 2 - Method 2. Map 1 – Method 2, Map 2 – Method 2.

The density maps derived from both methodologies for the 2015-2016 dataset manifest a congruence in spatial variations. The maps, shown in **Figure 9**, generated based on the total number of photons, consistently depict high-intensity regions within the x range of [-1000, 1000] arcseconds and the y range of [-500, 500] arcseconds. Notably, the apex of intensities is concentrated in query region three, corroborating the outcomes derived from intensity values. Despite this concordance in spatial variations, nuanced differences are discernible. The first method exhibits heightened intensity around the y range of [0, 500], while the second method showcases higher intensity within the range of [-250, 250].

Moreover, variations in intensity values are observable, with the first method portraying elevated intensities in regions two and one compared to the second method. Visualizations allow us to ascertain that the total intensities for both methods are comparable, with peak intensities for maps one and two reaching 6e-7 and 7e-7, respectively. These divergences in spatial patterns and intensity values underscore the intricate dynamics inherent in solar flare phenomena. The complexity of these dynamics necessitates a comprehensive understanding and careful consideration of methodological choices when analyzing and interpreting such intricate spatial variations and intensity distributions.

The visualizations using density maps for the 2004-2005 dataset demonstrated a similarity in spatial variations, where high-intensity regions were consistent between both methods. In contrast, the 2015-2016 dataset exhibited a variation in the query region with the highest intensities, highlighting differences in solar flare activity patterns between the two datasets. The overall lower intensities in the 2015-2016 dataset suggest a different solar flare behavior compared to the 2004-2005 dataset.

The main discrepancy between the two datasets lies in the query region with the highest intensities. While the 2004-2005 dataset consistently showcased the highest intensities in query region two, the 2015-2016 dataset exhibited the highest intensities in query region three. This shift suggests a temporal and possibly cyclical variation in the spatial distribution of solar flare intensities. Factors such as the solar cycle, magnetic activity, and other solar dynamics may contribute to these observed differences.

Overall, the computation of intensity values and the generation of density maps consistently yielded comparable results, indicating the efficacy of both methodologies in capturing trends related to solar flare occurrences and intensities across the solar surface. While the choice between methods may not significantly influence spatial distribution outcomes, the noted disparities between datasets emphasize the intricate and dynamic nature of solar activity. Variables such as solar cycle variability, instrumental distinctions, and the inherent complexity of solar flares contribute to these observed differences.

Despite these variations, the consistent findings across methods provide a degree of flexibility in choosing an appropriate methodology, enabling robust analyses even when a specific dataset might be unavailable for a particular method. This resilience in the face of dataset discrepancies enhances the practical utility of these methodologies for studying solar activity trends, facilitating a more comprehensive understanding of the nuanced factors influencing solar flare dynamics.

2.3 Spatiotemporal Analysis of Hotspot Discovery

In consideration of the outlined methodologies, it is imperative to acknowledge the nuanced strengths of Method 1 and Method 2. Method 1, focused on photon counts during a solar flare, provides granular insights into concentrated activities within specific regions. In contrast, Method 2 excels in portraying the comprehensive distribution of overall intensity.

To discover hot spots, we chose to go with Method 1. Our deliberate choice to employ Method 1 is underscored by its inherent alignment with the observational nature of solar flares, which are discerned through the photons emitted. The selection of Method 1, which centers on the total photon count during a solar flare, ensures a methodology that resonates with the fundamental principle that solar flares are observed through the photons released during these events.

To maintain simplicity, our analysis concentrated on one data set at a time over a two-year observational span. During the two-year observational span, dates were systematically categorized into eleven batches, each spanning four months with a two-month overlap with the preceding batch. Employing the two distinct thresholds, hotspots were identified, with the use of Kernel

Density Estimation to discern spatial concentrations of activity. Gaussian kernels, with a bandwidth of 250, were applied in conjunction with predefined thresholds to create density maps, facilitating the visualization and identification of hotspots. This process was iteratively conducted for each batch.

In the examination of solar flare events and hotspot discovery techniques, two intensity threshold values, denoted as D1 and D2, were instrumental in delineating and characterizing hotspots. D1 was tailored to pinpoint small yet intensely concentrated hotspots, whereas D2 was aimed to identify larger and more widespread hotspots. These hotspots were chosen visually as there was a max intensity of 8e-7 and a low of 0.

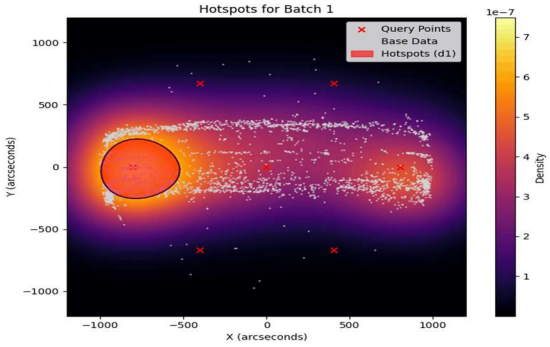


Figure 10: Time Series for hotspot estimation using threshold D1 for the 2004-05 dataset. (GIF)

In data from 2004-2005, employing the first threshold, D1, facilitated the generation of hotspot maps for each batch, thereby revealing distinct patterns. Significantly, batches two, three, eight, and nine displayed concentrated hotspots primarily within query regions two and three, with exceptions and variations attributable to temporal and regional dynamics. Conversely, batches five, six, and ten manifested hotspots confined to region three, indicative of potential heightened solar activity or complex flare events. As seen in **Figure 10**.

Interestingly, batches one, eleven, and four presented an anomaly as they showcased larger hotspots in region two, contrary to the intended purpose of the threshold, which was designed to pinpoint smaller, more concentrated spots. This incongruity prompts further scrutiny into the underlying factors contributing to the emergence of larger hotspots in region two during these specific batches.

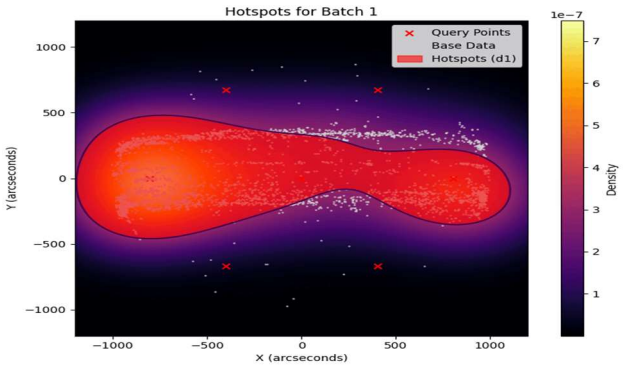
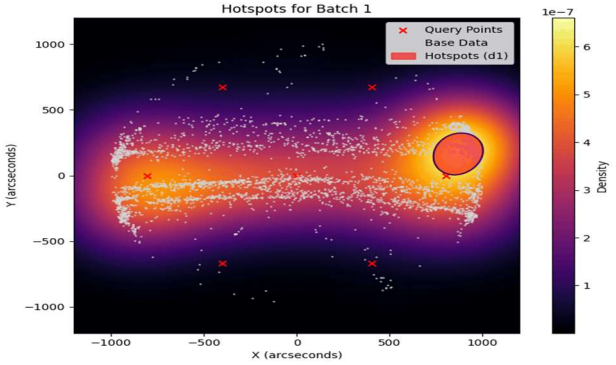


Figure 11: Time series for hotspot estimation using threshold D2 for the 2004-05 dataset. (GIF)

The application of the second threshold, D2, in generating hotspot maps across all batches revealed consistent and uniform results resembling an oval that is pinched in the middle. This characteristic shape indicates a concentration of hotspots within the range of regions one, two, and three, with the bulk of the activity centered in regions two and three. Although the size of each side of the oval varies across batches, the predominant trend is an emphasis on the left side of the domain or the left side of the Sun's surface in batches one through eleven. Noteworthy is batch five, where the distribution appears even but wider on the Y-axis on the right side in region three. (**Figure 11**)

A comprehensive examination of the hotspots over a time series facilitated a comparative analysis of maps generated in all batches. Within this temporal framework, significant differences in localization became apparent. While the application of threshold D1 did not reveal a continuous pattern in hotspot emergence, it did highlight the recurrence of highly intense hotspots around regions two and three. This observation finds support in the time series analysis of threshold D2, which, due to its continuous uniformity, indicates that hotspots generally tend to occur around the equator of the Sun.



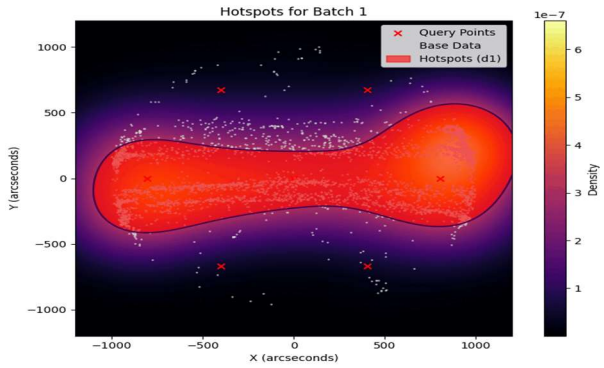


Figure 12.1 and 12.2: Time series using thresholds D1 and D2 (respectively) for the 2015-16 dataset. (GIF)

Building upon the analysis of the 2015-2016 data set, the observation of only one high-intensity hotspot in regions two or three during this period, as seen in the comparison of batches in D1, as shown in **Figure 12.1**, suggests a significant shift in solar activity dynamics. Additionally, the reduced banding across the equator in D2 comparisons further emphasizes the altered pattern during this timeframe, as seen in **Figure 12.2**.

The 11-year cycle, marked by the Sun's magnetic field flip and resulting in a solar minimum, introduces a unique phase in solar activity. As noted by the Space Weather Prediction Center, solar flares are closely associated with active regions characterized by strong magnetic fields. In light of this, the observation that regions two and three consistently exhibit high-intensity hotspots aligns with the notion that these regions may harbor strong magnetic fields.

The evolution of these magnetic fields over the solar cycle can play a pivotal role in influencing active regions on the Sun. The analysis suggests that the lack of recurring patterns over the two-year duration may be attributed to the dynamic nature of these magnetic fields during a solar maximum, influencing the emergence and localization of hotspots. The focus on regions two and three in D1 and the altered equatorial banding in D2 during the solar maximum period further supports the idea that magnetic field dynamics play a crucial role in shaping solar activity patterns.

The distinct areas of hotspots observed during the solar minimum in 2015-2016 underscore the sensitivity of solar activity to the Sun's magnetic field conditions. This insight contributes to a more comprehensive understanding of the variability in hotspot localization over different phases of the solar cycle. As such, the analysis not only enhances our understanding of solar flares but also highlights the importance of considering the broader solar context when interpreting observational data. Further investigations into the specific mechanisms through which magnetic fields influence hotspot emergence and distribution could provide valuable insights into the underlying processes driving solar activity variations.

3 Conclusion

In our comparative analysis of solar flare datasets from the Reuven Ramaty High Energy Solar Spectroscopic Imager, we explored temporal, spatial, and energetic dimensions of high-intensity solar flare events spanning the years 2004-2004 and 2015-2016. Our investigation aimed to deepen our understanding of solar flare dynamics, drawing insights from key variables such as date, duration, spatial position, and energetic characteristics.

Overall, our comprehensive analysis underscores the intricate and dynamic nature of solar flare phenomena. The observed variations between the datasets highlight the influence of solar cycles, magnetic field dynamics, and other inherent complexities. While Method 1 and Method 2 provided consistent findings, the dataset-dependent differences emphasize the need for careful consideration of methodological choices in solar activity analyses.

ACKNOWLEDGMENTS

Our team consists of Ly Ha, Erica Hay, Khuong Ngo, and Yasan Naser. Contributions will be written following each person's name.

Ly Ha: Code for Method 1 density maps, Code for hotspot discovery and time series for threshold 1 and 2, Spatiotemporal Analysis of Hotspot Discovery, Change Analysis of Hotspot Discovery.

Erica Hay: Code for Method 2 density maps, Analysis of Flare Intensity Estimation and Estimation Techniques, Analysis of duration of 2004-05 and 2015-16 datasets.

Khuong Ngo: Code for calculating basic statistics, Analysis of basic statistics in 2004-05 and 2015-16 datasets, Analysis of the Solar Flare Frequency Based on Monthly Distribution, Abstract of Helios Project

Yasan Naser: Task 1 Subtask 5, Task 2 Subtask f, g, and h. Unfortunately, did not participate in the completion of Task 3, this report.

REFERENCES

[1] NASA. (2022, June 10). Solar Flares FAQs. https://blogs.nasa.gov/solarcycle25/2022/06/10/solar-flares-faqs/#:~:text=The%20number%20of%20flares%20also,a%20few%20times%20per%20month.

[2] NASA. (n.d.). Solar Activity and Space Weather. The Sun's Magnetic Field. https://spaceplace.nasa.gov/solar-activity/en/#:~:text=The%20magnetic%20field%20lines%20near,energy%20called%20a%20solar%20flare.

[3] NASA. (n.d.). The Sun: By the Numbers. https://science.nasa.gov/sun/facts/

[4] NASA. (n.d.). What is a Solar Flare? https://www.nasa.gov/image-article/what-solar-flare/.

[5] OpenAI. (2022). GPT-3.5 "ChatGPT" Language Model. https://www.openai.com/research/chatgpt

[6] Space.com. (n.d.). New Predictions Suggest Solar Maximum for Solar Cycle 25 in 2024. https://www.space.com/solar-maximum-expected-2024-new-predictions-suggest

[7] UCAR Center for Science Education. (n.d.). Sun's Magnetic Field Rotate and Tangle (Movie). https://scied.ucar.edu/video/sun-magnetic-field-rotate-tangle-movie#:~:text=Over%20time%2C%20the%20Sun's%20differential,active%20regions%20where%20sunspots%20form.

[8] NASA. (n.d.). Solar Cycles: How the Sun Affects the Earth. https://spaceplace.nasa.gov/solar-cycles/en/#:~:text=The%20Sun's%20magnetic%20field%20goes%20through%20a%20cycle%2C%20called%20the,poles%20to%20flip%20back%20again