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

In this sample-structured document, neither the cross-linking of float elements and bibliography nor metadata/copyright information is available. The sample document is provided in “Draft” mode and to view it in the final layout format, applying the required template is essential with some standard steps.

These steps, which should require generation of the final output from the styled paper, are mentioned here in this paragraph. First, users have to run “Reference Numbering” from the “Reference Elements” menu; this is the first step to start the bibliography marking (it should be clicked while keeping the cursor at the beginning of the reference list). After the marking is complete, the reference element runs all the options under the “Cross Linking” menu.

For accuracy check of the structured paper, user can run the option **Manuscript Validation**. It informs the user of the wrong or missing values in the paper. The user must correct the paper as per validation messages and rerun **Manuscript Validation**.

Now, to generate the required layout of the paper, the user should select one of the template styles under the Define Template Style option (choose the required layout design, i.e. choose between Journals and Proceedings).

Some specific values are required to create a standard layout by choosing a template for the journals or proceedings. So once the user chooses one of the template layout styles, the respective Journal or Conference details dialog box (i.e. **journal/conference acronym, DOI, ISBN, copyright, year, etc.**) will appear as a prompt during the Define Template Style functionality. The user should fill these values, after which the template creates the desired layout of the paper. The user can now create a PDF of his/her manuscript using the “*Save as PDF*” option.

If the user is adding any new data, they should make sure to style it as per the instructions provided in previous sections. Carry out the steps for Cross-linking, Fundref data, adding Document History (specific to journal submission), and finally, Manuscript

CCS CONCEPTS

• Mathematics of computing • Probability and Statistics   •Multivariate statistics

KEYWORDS

Insert keyword text, Insert keyword text, Insert keyword text, Insert keyword text

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 Data from 2004-05 vs Data from 2015-16

The updated template, user manuals, samples, and required fonts, all are available at the URL <https://www.acm.org/publications/proceedings-template>. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in different clips. **Display Formula with Number**

 (1)

**Continuation part of Paragraph Text** The user must style this paragraph in **ParaContinue** style, which follows immediately after the **DisplayFormula** (numbered equation). The **DisplayFormula** style is applied only in case of a numbered equation. A numbered equation always has a number to its right. Insert paragraph text here. **Display Formula without Number**



The **DisplayFormulaUnnum** style is applied only in case of an unnumbered equation. An unnumbered display equation never contains an equation number to its right, and this unique property distinguishes it from a numbered equation.



Figure 1: Figure Caption and Image above the caption [In draft mode, Image will not appear on the screen]

**Theorem/Proof/Lemma.** Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement.

....Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract.

1.1 Heading Level 2

In the below paragraph, it is explained how alt-txt value is placed in **MS Word 2010**. To add alternative text to a picture in Word 2010, follow these steps:

1. In a Word 2010 document, insert a picture.
2. Right click on the inserted picture and select the **Format Picture** option.
3. Select the **Alt Txt** option from the left-side panel options.
4. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

Below are steps to place alt-txt value in **MS Word 2013/2016**. To add alternative text to a picture in Word 2013/2016, follow these steps:

1. In a Word 2013/2016 document, insert a picture.
2. Right click on the inserted picture and select the **Format Picture** option.
3. In the settings at the right side of the window, click on the "Layout & Properties" icon (3rd option).
4. Expand **Alt Txt** option.
5. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

*1.1.1 Heading Level 3.* Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

*1.1.1.1 Heading Level 4.*Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

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 #.

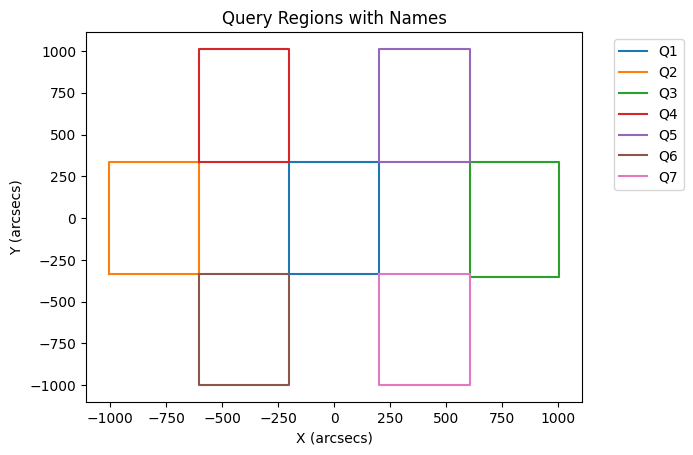
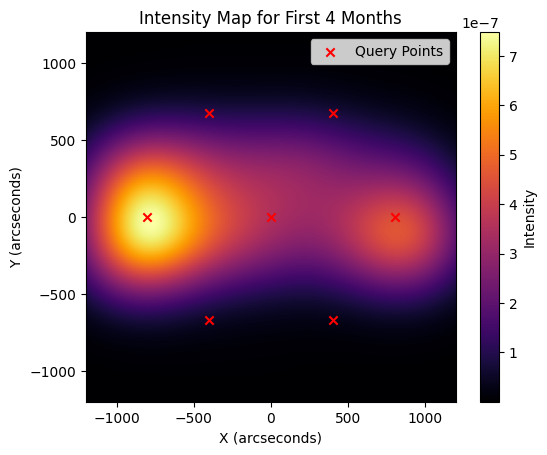
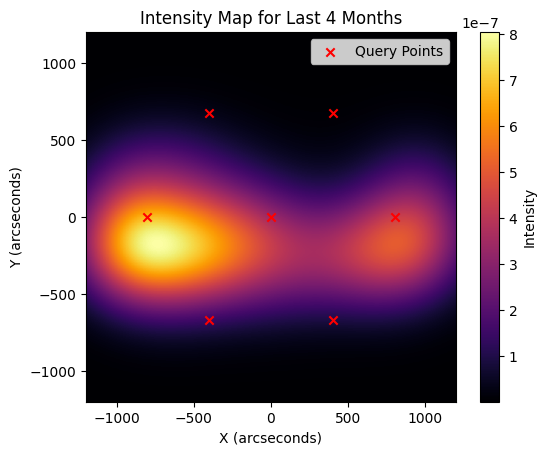


Figure #: Query Regions with names, as specified above.

In the realm of Method 1, we gauged the intensity of solar flare events in distinct sun regions using kernel density estimation with a 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. 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.

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.

A chart with a colorful background

Description automatically generated with medium confidenceWhen 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.

A chart of a galaxy

Description automatically generated with medium confidence

Figure #: 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 #). 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. 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]. 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.

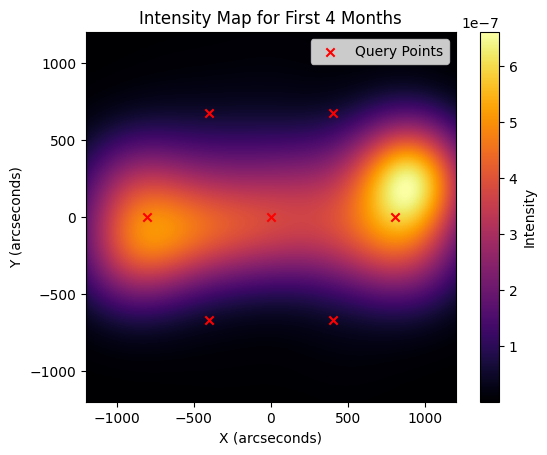
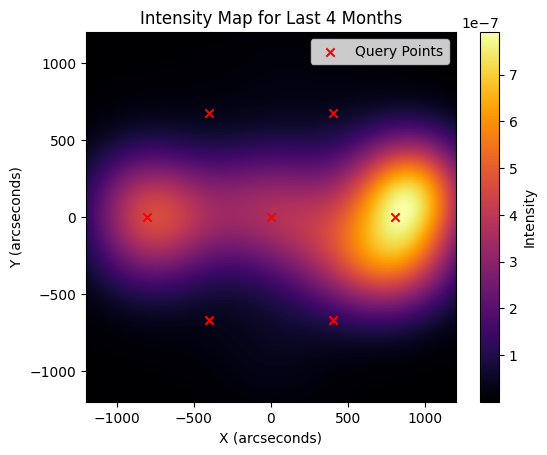
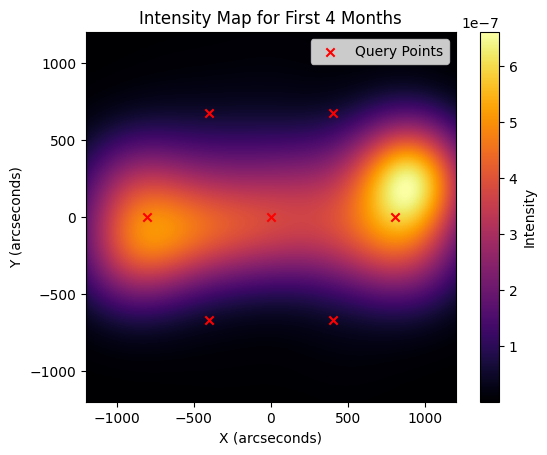
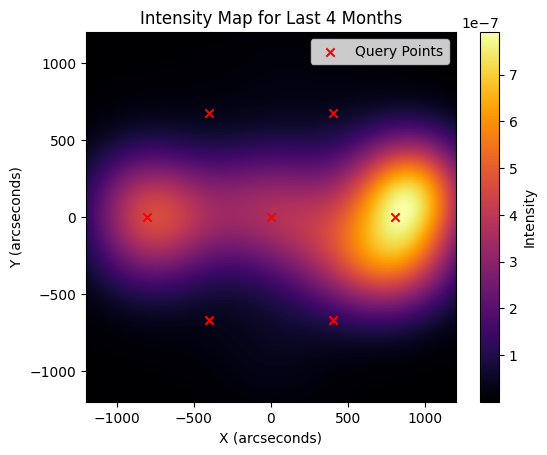


Figure #: Density/Intensity Maps for 2015-16 Dataset in order from left to right. Map 1 – Method 1, Map 2 - Method 2. Map 1 – Method 2, Map 2 – Method 2.

The density maps derived from both methodologies for the 2015-2016 dataset (Figure #) manifest a congruence in spatial variations. The maps, 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.

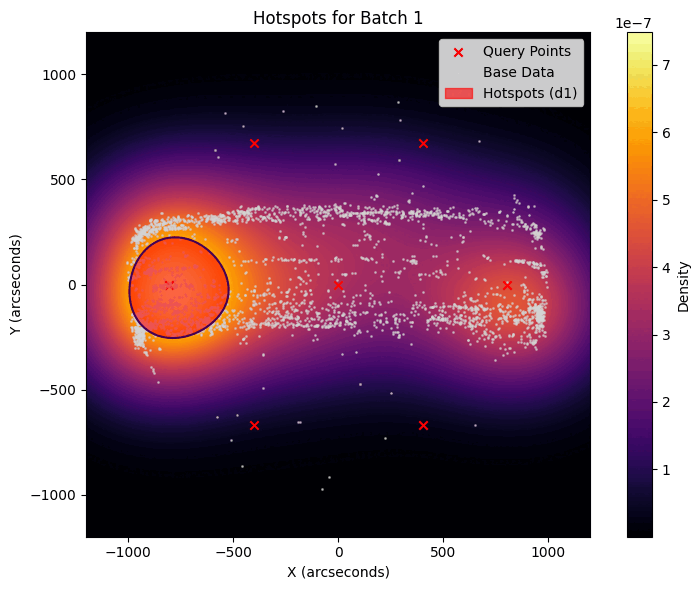


Figure #: Time Series for hotspot estimation using threshold D1 for the 2004-05 dataset.

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 #.

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.

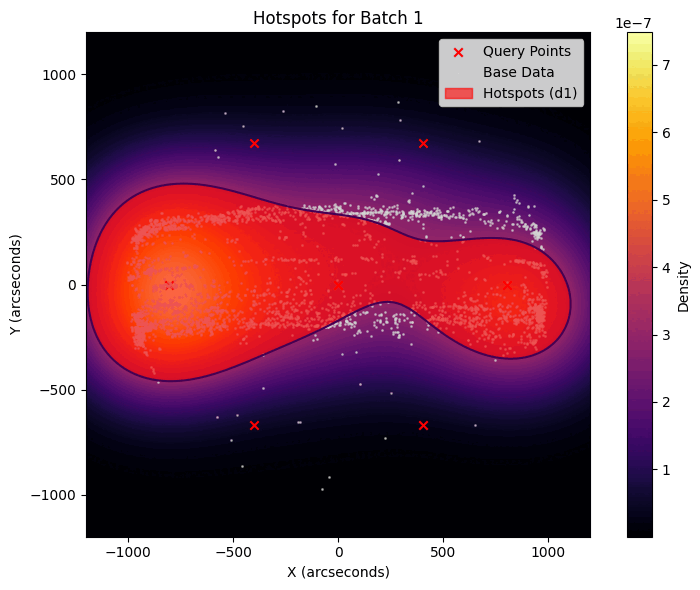
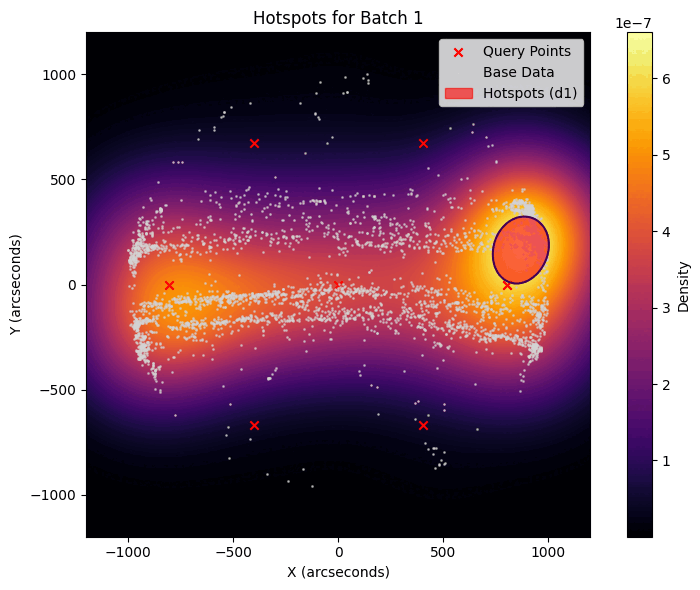


Figure #: Time series for hotspot estimation using threshold D2 for the 2004-05 dataset.

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 #)

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.



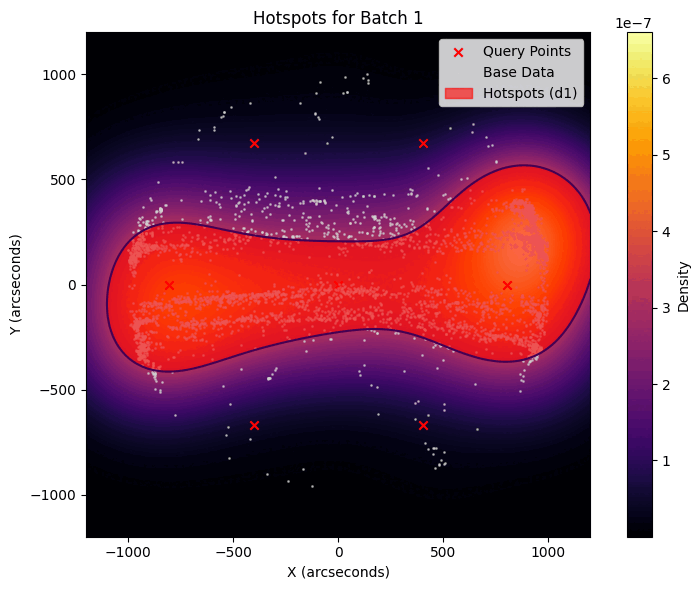


Figure #: Time series using thresholds D1 and D2 (respectively) for the 2015-16 dataset.

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, 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. (Figure #)

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.

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.

Erica Hay: Code for Method 2 density maps, Analysis of Flare Intensity Estimation, 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.

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] 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

[6] 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.

[7] 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.

Conference Name:ACM Woodstock conference

Conference Short Name:WOODSTOCK’18

Conference Location:El Paso, Texas USA

ISBN:978-1-4503-0000-0/18/06

Year:2018

Date:June

Copyright Year:2018

Copyright Statement:rightsretained

DOI:10.1145/1234567890

RRH: F. Surname et al.

Price:$15.00