Helios Project

EDWIN PALACIO (1333722), MOISES PALACIOS(1920759), KLEIN HOUMANI (2184415), ABEL BARCENAS(1919301)

TASK 1: SOLAR FLARE INTENSITY ESTIMATION

Our analysis centers on segmenting and assessing solar flare data spanning 2004 to 2005 into methodically overlapping four-month intervals. We introduce two innovative estimation methods to evaluate flare intensities at specific coordinates, utilizing different solar flare attributes for a more comprehensive analysis. These methods are further explored to generate intensity maps at two distinct stages of the data set, facilitating a comparative study of the spatial and intensity variations of solar activity over time.

Method 1: Intensity Based on total.counts

Method 1 relies on the "total.counts" variable to establish intensity levels. Observations are categorized into low, medium, and high based on the distribution of the data, with thresholds determined by the density of total.counts.

Method 2: Intensity Based on duration.s and energy.kev

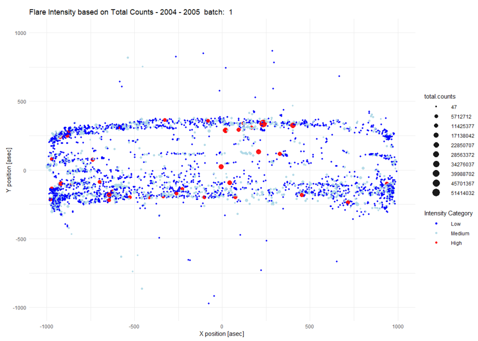
Method 2 calculates intensity through integration over a specified duration, prioritizing the contribution of "energy.kev" over "duration.s." This method dictates the size and color of observations in the heatmap, with color corresponding to six energy band categories.

Visualization of Heat Maps

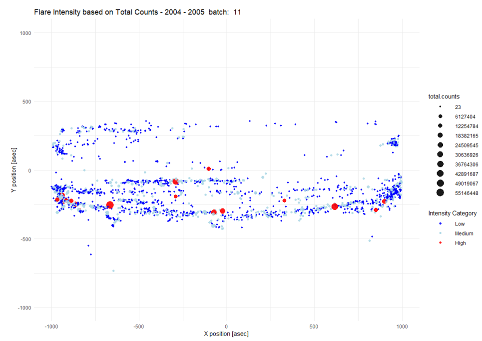
The dataset is partitioned into eleven four-month batches with a two-month overlap, facilitating detailed temporal analysis. For example, batch 1 includes months 1, 2, 3, and 4, while batch 11 includes months 21, 22, 23, and 24.

Author's address:
Edwin Palacio (1333722), Moises Palacios(1920759), Klein Houmani (2184415), Abel Barcenas(1919301).

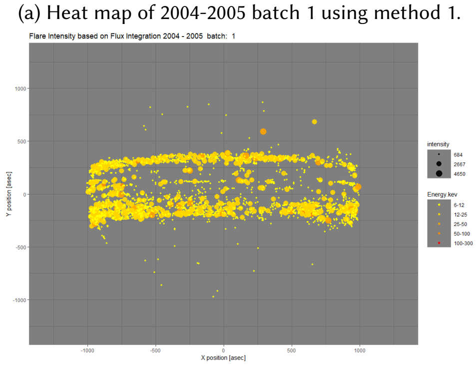
Manuscript submitted to ACM



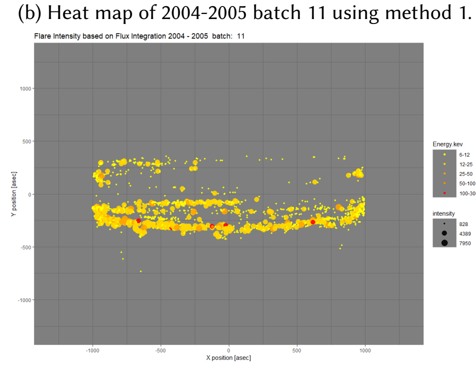
(a) Heat map of 2004-2005 batch 1 using method 1.



(b) Heat map of 2004-2005 batch 11 using method 1.



(c) Heat map of 2004-2005 batch 1 using method 2.



(d) Heat map of 2004-2005 batch 11 using method 2.

Fig. 1. Comparative heat maps for batches 1 and 11 using methods 1 and 2.

Comparative Analysis

Using Method 1, The generated heatmaps for the first four months (batch 1) of 2004 and the last four months (batch 11) of 2005 reveal a concentration of solar activity near the sun's center, predominantly between y=-500 and y=500. In both sets, the southern portion of the sun below y=0 exhibits the highest activity, indicated by dense clusters of points. However, batch 1 displays a more even distribution between the northern and southern halves, while batch 11 shows significantly less solar activity in its northern half. Analysis based on the covered area suggests higher overall solar activity in the first four months of the year compared to the last four months.

Applying Method 2 allows visualization of the energy bands associated with each observation. Batch 11 exhibits a greater occurrence of solar flares in the 100-300 energy band compared to batch 1. Interestingly, despite batch 1 having higher overall solar activity based on total intensity calculated by both methods, batch 11 showcases more solar flares of higher-level energy bands.

TASK 2: HOTSPOT DISCOVERY AND ANALYSIS

Hotspot Discovery Algorithm and Thresholds

To identify hotspots, we created a function named visualize hotspots, utilizing the stat density 2d function. We set our threshold values, d1, and d2, at the 95th and 75th percentiles, respectively. When invoking the visualize hotspots function for each threshold, data points falling below the respective threshold are omitted from the dataset. Additionally, the animation library is used afterward to create the GIFs.

Time Series Summarization for d1 Threshold

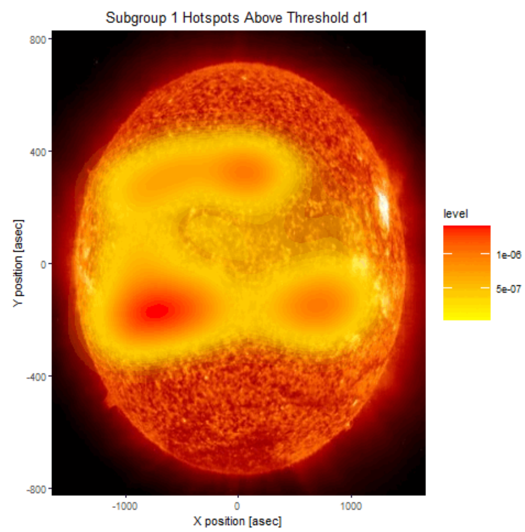


Fig. 2. GIF of hot spot visualization using threshold D1 using 95 percent quantile. Background images were taken within the time period of their respective subgroup. Images were provided from the Solar and Heliospheric Observatory (SOHO).

Our time series analysis of solar flare hotspots using the d1 intensity threshold reveals a distinct pattern of solar activity. The distribution of hotspots is non-uniform, with a tendency to cluster in certain active regions that recur across the observed batches. This clustering suggests that these regions are the epicenters of solar flare activity. Throughout the 11 batches representing different time periods, we see a clear variation in flare intensity. Some intervals exhibit densely packed hotspots with high intensity, potentially corresponding to a solar maximum, while others show sparse and less intense activity, aligning with what might be a solar minimum. Spatially, the activity concentrates in specific areas of the solar disk, suggesting a correlation with the sun's magnetic structures that govern flare occurrence. Temporal shifts in hotspot locations are also evident, with some periods showing significant activity near the solar equator and others displaying it closer to the poles.

Time Series Summarization for d2 Threshold

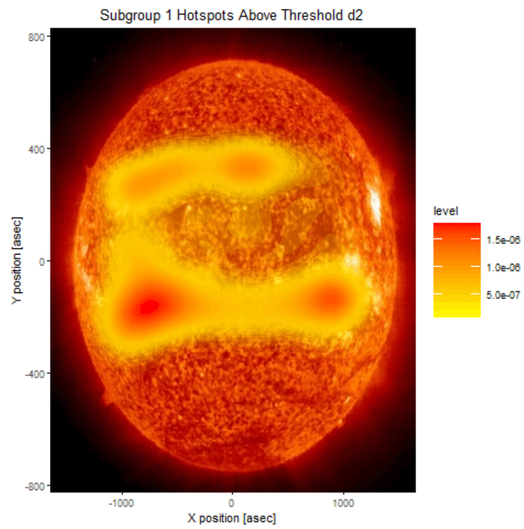


Fig. 3. GIF of hot spot visualization using threshold D2 using 75 percent quantile. Background images were taken within the time period of their respective subgroup. Images were provided from the Solar and Heliospheric Observatory (SOHO).

The time series analysis of solar flare hotspots using the d2 intensity threshold across the 2004-2005 period reveals a consistent pattern of solar activity. Notably, the hotspots, indicative of moderate to high-intensity flaring, predominantly occupy the edges of the solar disc. This pattern persists across the batches, with the central regions consistently exhibiting less activity, manifesting as either yellow or uncolored areas. The recurring appearance of the most intense flares—marked in red—at the edges suggests a propensity for these events to arise in the peripheral zones of the Sun. This could be reflective of the magnetic field lines' distribution, which are known to be more complex and potentially more unstable at the solar surface's outer reaches. Throughout the sequence of batches, we observe that the intensity of the peripheral hotspots varies. Some batches present with tightly concentrated red zones, signaling episodes of heightened flare activity, which could correspond to an active phase in the solar cycle. In contrast, other periods show these red zones as less pronounced and more scattered, pointing to a decline in activity, potentially indicative of a quieter solar phase.

Comparison of Time Series for d1 and d2 Thresholds

The time series with the d2 threshold mostly follows a similar pattern to the time series with the d1 threshold. Both share hotspots in similar areas, likely due to them sharing a lot of data points. However, for a lot of these hotspots, the density relative to the total amount of data points is lower when using the d2 threshold when compared to using the d1 threshold. This suggests that the higher intensity points tend to cluster together more and are largely contributing to the formation of hotspots. Once the threshold is lowered, the newly introduced lower intensity points tend to not cluster as much as the higher intensity points. This results in the relative density of the hotspots containing the higher intensity points being lowered.

TASK 3: COMPARATIVE ANALYSIS OF 2004-2005 AND 2015-2016 SOLAR FLARE DATA

Mean Intensity of Solar Flares

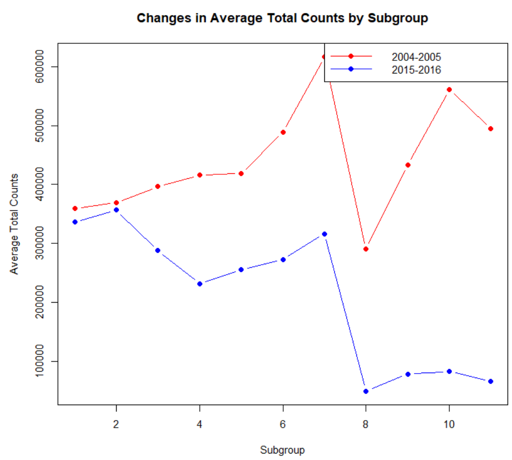


Fig. 4. plot of batches v average total count

Based on the heat map provided by method 1, Set 1 (2004-2005) exhibits consistently higher average total counts of solar flare activity across all subgroups compared to Set 2 (2015-2016), suggesting a period of heightened solar activity in the year 2004. By plotting the average total count by batch for each year a similar pattern between the two emerges despite a lower overall total count average in 2015. Solar flare activity seems to rise and fall during the same months the only difference is the intensity level. Given that the both sets of data come from a similar period of an 11 year solar cycle, this reinforces the idea that this solar cycle follows a cyclical pattern in regards to solar flare activity.

Variance in Solar Flare Intensity

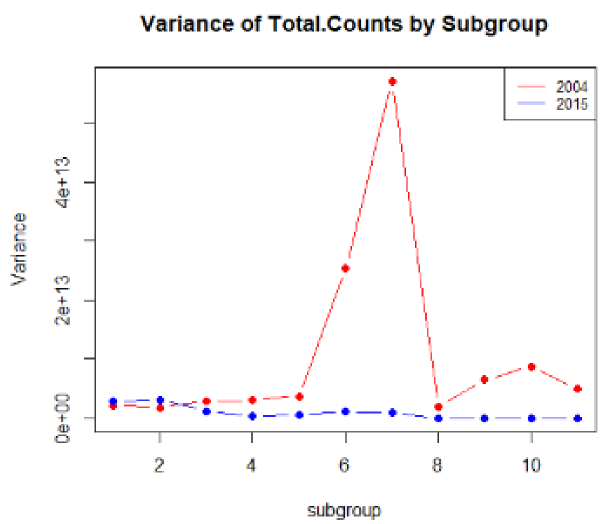


Fig. 5. variance of total.counts by batches

When observing the variance for the two different sets, there is a very significant spike in variance for subgroups 6 and 7 in Set 1. This suggests that these two subgroups contain more solar flares with extreme intensities, with subgroup 7 having the most extreme ones. Additionally, subgroups 3 - 11 in set 1 all had a higher variance than the same subgroups in set 2. The Set 2 data remained relatively consistent across all subgroups when compared to Set 1, having no obvious peaks.

Range of Solar Flare Intensity

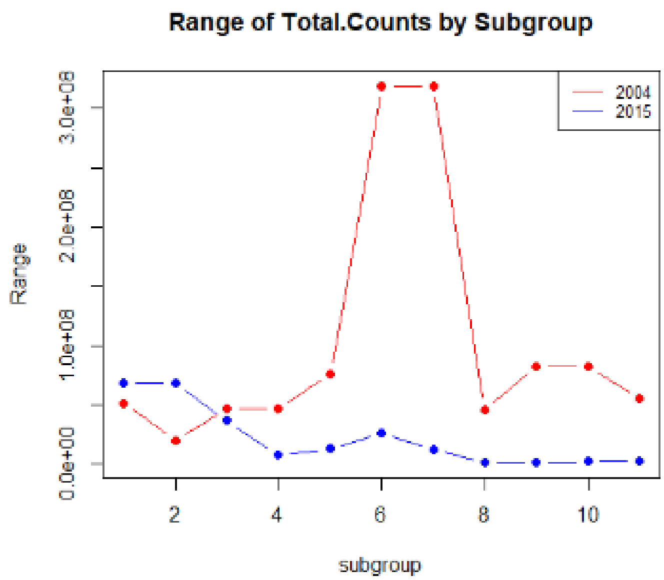


Fig. 6. range of total.counts by batches

Similar to what was observed in the variances, subgroups 6 and 7 in Set 1 have a significantly greater range than that of the other subgroups. Again this suggests that these two subgroups experienced a greater amount of solar flares with more extreme intensities. Subgroups 3 - 11 in set 1 also saw a greater range than the same subgroups in Set 2. Set 2 has a slight downward trend, suggesting that the intensity of solar flares may tend to be less extreme towards the end of the solar cycle.

Duration of Solar Flares by Subgroup

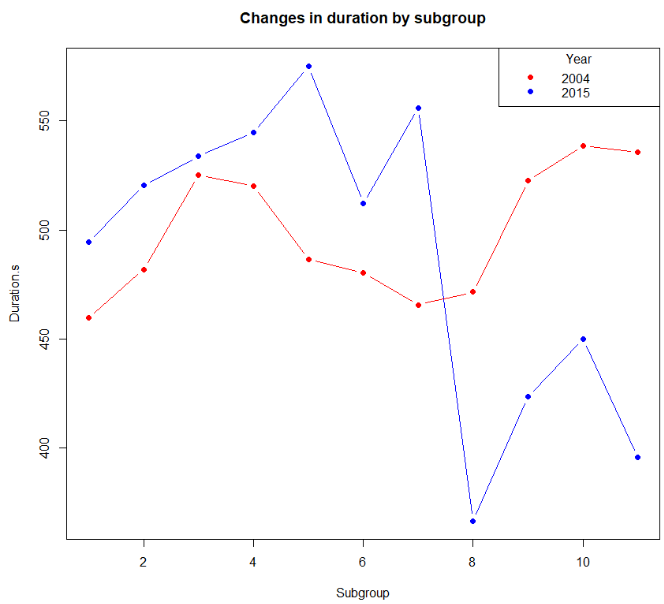


Fig. 7. duration by batches

The 2015-2016 dataset has longer durations for the first 7 batches. After the 7 batches the roles are reversed where 2004-2005 batches seem to have significantly longer durations now. This point is a potential factor on why method 2 has a much higher frequency of higher energy.kev values on dataset 2015-2016 which are visualized by orange and red. It can be seen that the 2015-2016 dataset provides longer and more intense solar flares during the first year and drops when beginning 2016. While 2004-2005 kept this consistent intense flares throughout the 2 years. With these observations I can see that there is a relationship between longer solar flares and higher intensities.

Energy Band Distribution

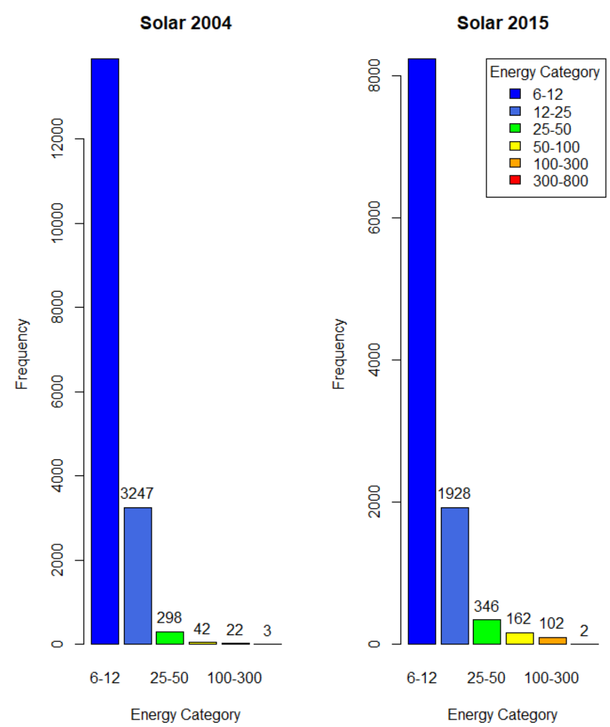


Fig. 8. energy band based on batches

Despite having fewer total observations, the 2015-2016 dataset reveals a greater proportion of solar flares within the higher energy bands when compared to the 2004-2005 data, a trend that is clearly illustrated by our two graphical representations. This suggests that while solar flare events were less frequent in the later period, they were more likely to be of higher energy when they did occur.

Spacial Variation:

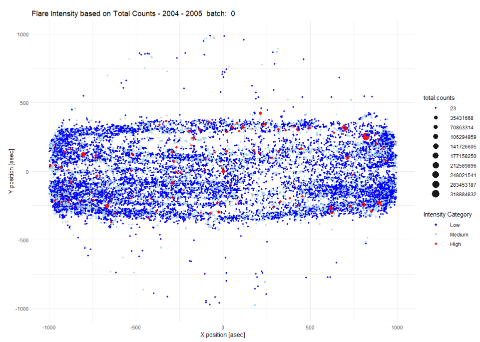


Fig. 9. method 1 using the entirety of 2004-2005

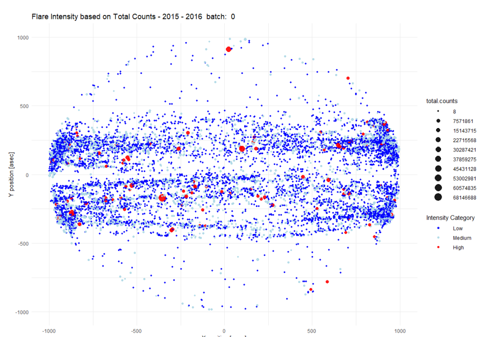


Fig. 10. method 1 using the entirety of 2015-2016

The two graphs, representing total.count for the years 2004-2005 and 2015-2016, display distinct spatial distributions of solar activity. The 2004-2005 graph shows a higher concentration of flares, with more frequent and intense activity, especially around the center of the Sun. On the other hand, the 2015-2016 graph reveals flares that are less densely packed and more evenly spread across the Sun's surface, suggesting a period of reduced solar activity. The clear contrast in how the flares are spread out over the Sun in these two datasets.

Conclusion

Across these statistics, the mean, variance, and range all tend to suggest that set 1 has a higher level of solar flare activity compared to set 2. This can primarily be seen in subgroups 6 and 7 in set 1, which consistently saw higher values for each of these statistics. Additionally, set 2 appeared to be more consistent relative to set 1. Across all the statistics, it did not appear to have any major peaks like set 1 has. Also, the average total counts across both sets followed a relatively consistent pattern, reinforcing the idea that this solar cycle follows a cyclical pattern in regards to solar flare activity.

CONTRIBUTIONS

Edwin Palacio: Came up with a method to calculate intensity using duration and energy.kev for task 1 method 2. Wrote code for visualization of task 1 and wrote an interpretation of heatmap generated by both method 1 and method 2. Generated graphs for task 3 for average total counts and distribution of energy bands category

Moises Palacios: implemented code for method 1. provided and explained simple basic statistics for energy band distribution, duration by subgroup, and average total count for task 3. In charge of formatting the report.

Klein Houmani: Wrote the code, created graphs, explained the algorithm used, and wrote the comparisons for task 2. Created the graphs and wrote explanations for the variance, and range in the basic statistics section of task 3. Organized our code together and wrote the October 27th status update.

Abel Barcenas: Assisted in the summarization and interpretation of time series data for task 1. Wrote the overall conclusion for the report, encapsulating the study's key findings and their significance.

REFERENCES

[1] ChatGPT. (2023). Assistance with LaTeX document formatting and R code troubleshooting [Chatbot interaction]. OpenAI.
[2] SOHO (Solar and Heliospheric Observatory). (n.d.). EIT: Extreme ultraviolet Imaging Telescope. Retrieved from https://soho.nascom.nasa.gov/data/summary/eit/
[3] LMSAL (Lockheed Martin Solar and Astrophysics Lab). (n.d.). SolarSoft Latest Events Archive. Retrieved from https://www.lmsal.com/solarsoft/latest_events_archive.html

In our analysis, images of the Sun for the months in the dataset were obtained from the Solar and Heliospheric Observatory (SOHO) [1] and the SolarSoft Latest Events Archive [2]