Summarization, Mapping, Hotspot Discovery, and Change Analysis of High-Intensity Solar Flare Events

Exploring Solar Flares with Data Science

Rayyan M. Rahman
University of Houston
Houston TX USA
rmrahman@CougarNet.uh.edu

Timothy Chou
University of Houston
Houston TX USA
thchou@CougarNet.uh.edu

Ryan Ball
University of Houston
Houston TX USA
rwball@CougarNet.uh.edu

ABSTRACT

The Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI), a NASA solar flare observatory launched in 2002, has been instrumental in exploring the physics of particle acceleration and energy release during solar flares. This research seeks to delve into the unique solar events that unleash substantial energy stored in the solar atmosphere. A specific focus is placed on understanding the mechanisms behind solar flares, which emit X-rays and longer-wavelength emissions, illuminating the impacts of these high-intensity solar activities on space-based observations.

The dataset under analysis is segmented into two distinct sets covering the time frames of 2004 to 2005 and 2015 to 2016. These carefully chosen intervals are aligned with the solar cycle, allowing for a comparative analysis while ensuring the data's relevance. The data cleansing process included the removal of the 3-6 KeV band and radial outliers to refine the dataset, preserving the accuracy of the spatial and energy-related information. The coordinate system for the dataset is structured in arc-seconds, offering detailed insight into the specific locations, highest energy levels, and peak counts of events. It's important to note that the solar flares' peak emissions were identified within the specified energy band, such as the 12 KeV to 25 KeV range.

Plotting this dataset without any transformation reveals intriguing spatial patterns, considering the Sun's center (0,0) in the coordinate system and recognizing its diameter as 0.5 degrees, equivalent to 30 arcminutes or 1800 arcseconds. The X and Y axes illustrate the horizontal and vertical dimensions, respectively, within the solar flare dataset. Notably, observations in 2004-2005 indicate a concentration of events near the Sun's equator, whereas the 2015-2016 dataset portrays a shift with intensities observed towards the Sun's poles. These spatial disparities prompt further investigation to decipher potential implications on solar activities during the specific solar cycles.

1 Solar Flare Intensity Estimation

Our first task involves subdividing Set 1 (2004 - 2005) into smaller 4-month batches with a 2-month overlap, creating eleven batches from months 1 to 24. This will be used again in our next task as well. We also needed to develop two distinct methods of estimating flare intensity. Method 1 assesses intensity based on the "total.counts" attribute, while Method 2 gauges intensity using attributes "duration.s" and "energy.kev." Intensity maps for the initial 4 months (months 1+2+3+4) are created using both Method 1 and Method 2, as well as for the final 4 months (months 21+22+23+24). These maps compare spatial variation and total intensity.

1.1 Method One

For method one, we decided to scale the attributes. The scaling process typically involves transforming the data to have a mean of 0 and a standard deviation of 1. This normalization can aid in making different datasets or variables comparable, particularly when they have different ranges or units. The primary attribute for this method is the ‘total.counts’ attribute that specifies that total count in corrected counts where counts are in the energy range of 6-12 kEV integrated over duration of flare summed over all sub collimators, including background. By scaling the 'total.counts' attribute in the context of analyzing solar flare events, it's likely aimed at providing a more standardized measure of intensity that's consistent across the dataset. This can assist in meaningful comparisons and analyses, especially when dealing with solar event intensities that might span a wide range of values.

1.2 Method Two

Method two is designed to calculate intensity values for each entry in the dataset by combining the 'total.counts' and 'energy.kev' attributes. The 'energy.kev' attribute is the highest energy band in which the flare was observed. There are a total of 8 bands that can be observed starting from 6 keV to 20000 keV. This process involves assigning specific weights to different energy bands and then multiplying these weights by the total counts. The weights assigned to each energy band provide a customized scale for assessing the intensity, allowing for a more tailored and nuanced evaluation of the solar flare events. This approach assumes that the intensity of a solar flare event is influenced by both the energy range and the total counts within that range.

1.3 Intensity Analysis

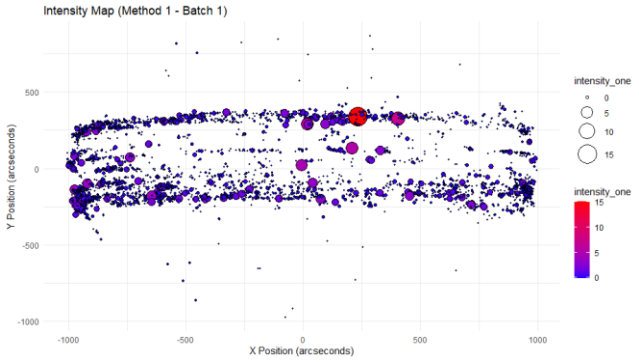

Figure 1: **Intensity map using method one for the first batch from set 1 from Set 1 (2005 – 2006)**

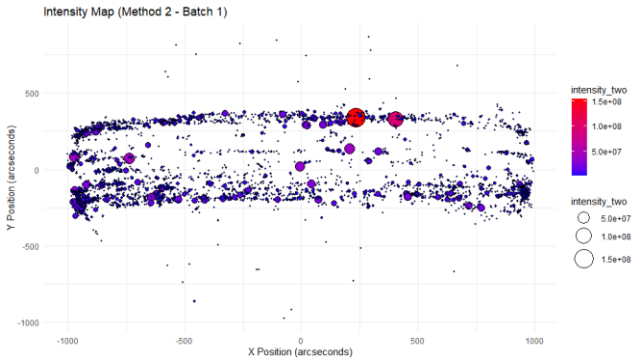

Figure 2: **Intensity map using method two on the eleventh batch from Set 1 (2005 – 2006)**

Spatial variation between the batch 1 intensity maps is very similar. Clusters are present in Batch 1 from x = -1000 to -800 and y = -270 to -100. The higher intensity points are present in the same locations across batch 1. We notice that most of the data is concentrated between y = -270 to 270. The total intensity across both maps is vastly different as we can see by the key on the right. However, the scaling of the intensities across method 1 and 2 produces similar spatial variation and spacing across the batch 1 maps.

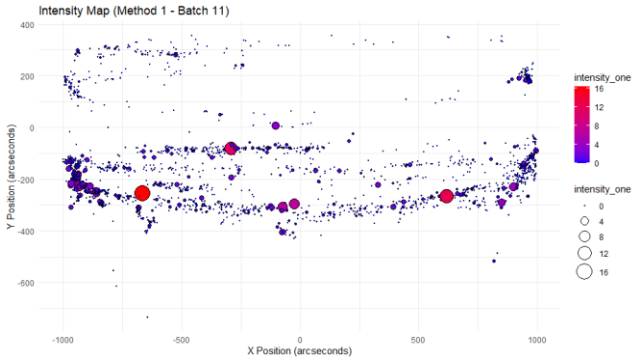

Figure 3: **Intensity map using method two on the eleventh batch from Set 1 (2005 – 2006)**

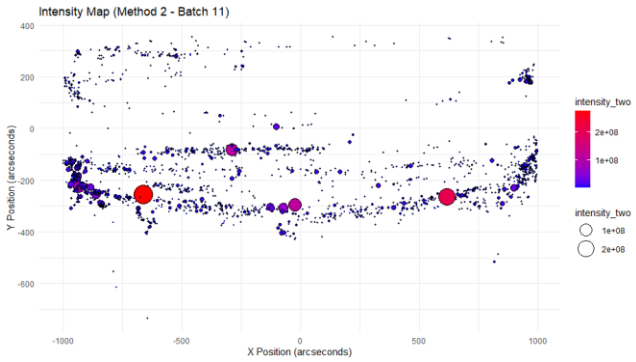

Figure 4: **Intensity map using method two on the eleventh batch from Set 1 (2005 – 2006)**

For batch 11, we also see that both intensity maps are similar. Especially when it comes to spatial variation where the points across both maps are similar. We see clustering around x = -1000 to -800 and y = -270 to -250. Total intensities is the same scenario with batch 1 where the scaling evens out on the plot.

2 Hotspot Discovery Analysis

The hotspot discovery algorithm is designed to identify distinct and contiguous areas of increased solar flare intensity within the 2D X-Y space, a vital aspect in the study of solar phenomena. Following the design, the next crucial steps involve the determination of two key intensity thresholds, d1 and d2, essential for Method one's hotspot discovery. Based on the data, we decided to set the d1 threshold to 0.85 and d2 threshold to 0.5. After creating a visualization technique in R using the 2d density map we then created time series data for the 11 batches using the intensity thresholds d1 and d2, thus generating a comprehensive view of hotspot dynamics across multiple temporal segments.

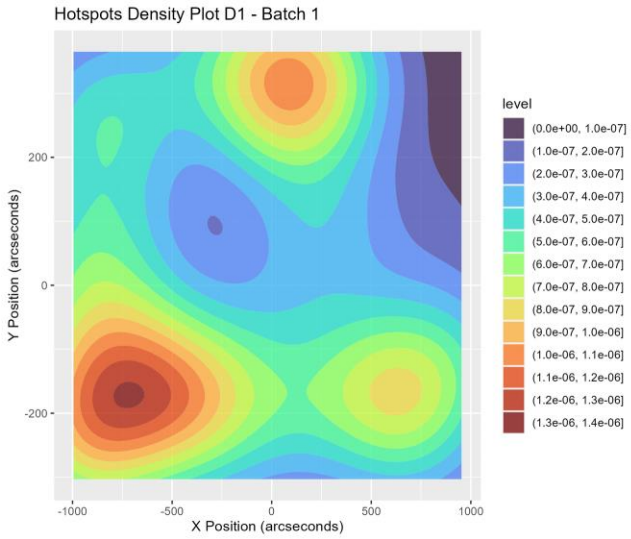

Figure 5: **Timeseries density hotspot map using d1 threshold**

The observation of hotspots at specific locations (around x = ~800 and y = ~-200) in four different batches and the consistent tendency for the hottest spots to occur towards the edges but not in the middle indicates an interesting pattern. The recurring high-intensity events at those specific spatial coordinates suggest a localized pattern or a higher likelihood of event occurrence in those areas across different periods (batches). Additionally, the consistent absence of such high-intensity events in the middle areas implies a potential spatial bias in the distribution of these events, favoring certain peripheral regions over central regions.

Most of the hotspots occurring below the y = 0 could indicate a few potential factors. Hotspots tend to concentrate at mid-latitudes or higher solar latitudes, moving towards the poles as the solar cycle progresses. This migration is a typical characteristic of the Sun's 11-year activity cycle. The uneven distribution might be correlated with the current phase of the solar cycle. During different stages of the solar cycle, the distribution of sunspots and solar activity varies, often with a higher concentration away from the solar equator.

In context these hotspot indicators in a grand scope on happening in the middle of the sun. If we compare the intensity maps from earlier to these, you will notice that the x arc coordinates have a wider range compared to the hotspot plots.

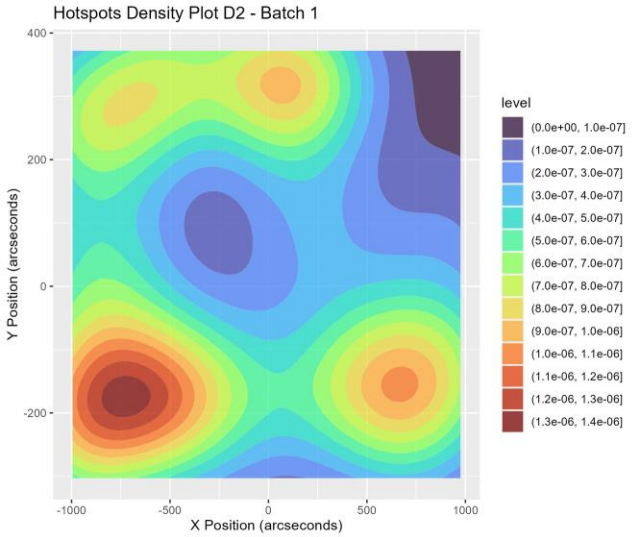

Figure 6: **Timeseries density hotspot map using d2 threshold**

The same pattern of hotspot distribution is observed for threshold d2 as for d1, it could imply that both thresholds are identifying regions with high solar activity, albeit possibly with varying intensity criteria. The threshold is slightly lower for d2 (0.5) vs d1 (0.85) does show more hotspot activity because we are including more data points to be mapped. These hotspots are still being mapped generally towards the edges of the sun's surface. However, this can also show that there is some hotspot activity towards the center but not enough to recognize a pattern across the 11 batches.

Comparing the plots from D1 & D2 thresholds shows that the pattern consistency across thresholds might further reinforce the significance of the latitudinal and longitudinal positioning of these active regions. The prevalence of hotspots in specific areas, particularly if they are consistent between different intensity thresholds, signifies certain stability in the behavior of these solar regions. It also supports the concept that the

hotspots are aligned with specific features in the Sun's magnetic field, potentially associated with the magnetic sunspot cycle

3 Change Analysis for Solar Flares

In Task 3, the objectives were to summarize the differences between the dataset 1 (2005 - 2006) and dataset 2 (2015 - 2016). On top of summarizing the differences, Compare the two kinds of intensities you analyzed before and also analyze spatial variation.

3.1 Spatial Variation Analysis

Comparing Figures 1 & 2 from Set 1 to Figures 7 & 8 from Set 2 we see they have polarizing differences. The first observation is that Set 1 has primarily 2 distinct bands of intensities that sit equal distances away from each other centered on the Sun's equator. Set 2 has 3 bands of intensity with the first band sitting above the equator and the lower 2 bands being below the equator close in proximity. Additional observations between the sets that are 10 years apart include the focus of intensities have shifted from sitting on top of the outer bounds of the bands to being more evenly distributed along the whole band.

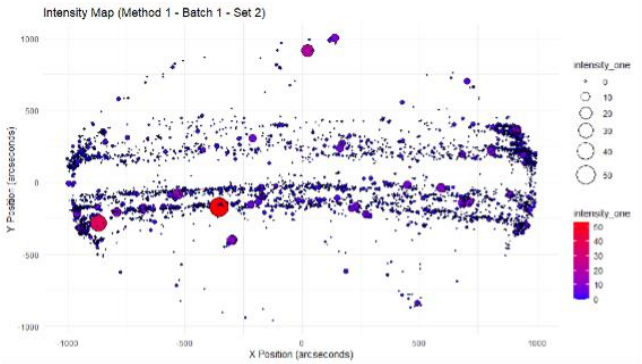

Figure 7: **Intensity map using method one on the first batch from Set 2 (2015 – 2016)**

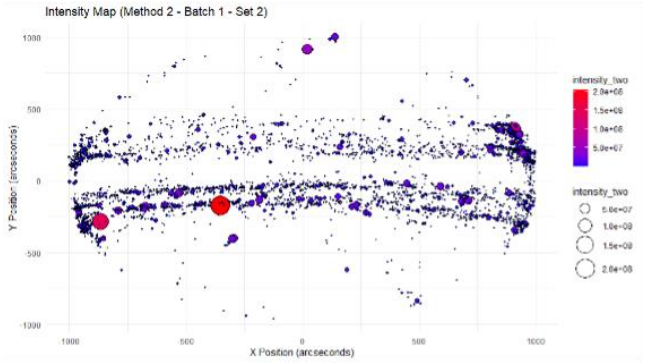

Figure 8: **Intensity map using method two on the first batch from Set 2 (2015 – 2016)**

Changes in Set 2 from 2015-2016 visually appear to have fewer intensities than Set 1. There are increased large intensities along the poles of the Sun in comparison to Set 1. This could indicate a shifting of intensities towards the poles as time elapses. This could be due to the phenomenon that we discussed early about the magnetic field structure and the presence of sunspots. Solar activity is associated with sunspots, which are relatively cooler and darker regions due to intense magnetic activity. Sunspots tend to occur closer to the equator during the sun's peak activity and closer to the poles during the quieter phases. An additional observation that can be made is that in method 2, we see higher intensities at x = 500 and y = 900. Overall, the results are similar between both methods.

When comparing batches eleven for methods one and two (figures 9 and 10), the intensities have once again appeared to increase between the two methods. This change can be observed at coordinates x = -200 and y = -750.

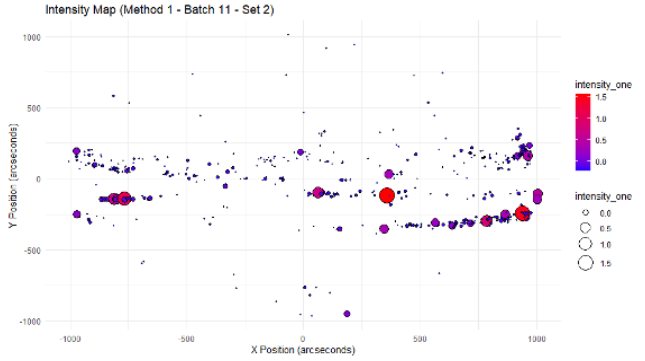

Figure 9: **Intensity map using method one on the eleventh batch from Set 2 (2015 – 2016)**

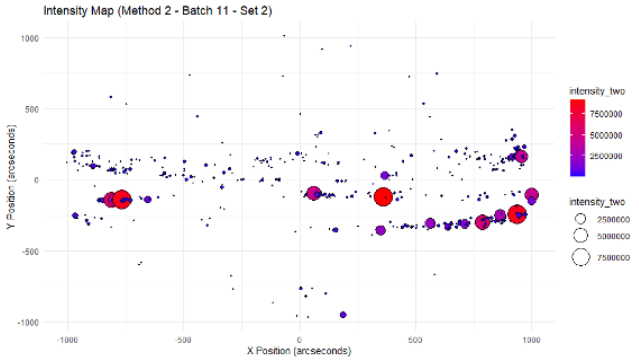

Figure 10: **Intensity map using method two on the eleventh batch from Set 2 (2015 – 2016)**

When comparing the differences for batches one and eleven for method one, the distribution of intensities are much less uniform and take on a sparser approach. The largest of the intensities no longer appear below the equator on the left quadrant but now appear below the equator in the right quadrant at coordinates x = -250 and y values 500 to 1000. This shift also includes a similar number of larger intensities that are slightly smaller.

The differences between batches one and eleven for method two are easier to identify visually in the scatter plots. A similar movement is witnessed where distribution is less uniform and sparser. It is also witnessed that the larger intensities perform a similar shift as method 1 with similar coordinate positioning. The only significant change that is witnessed is the size of the intensities observed relative to method one in that quadrant. We see that larger intensities are now more clustered and larger than their method 1 counterparts.

The overall changes for both methods one and two between batches one and eleven for Set 2 is that there has been a shift from a uniform distribution along the equator to below the equator. This shift also has the largest amount of intensities clustering in the bottom left quadrant in both scenarios. Additionally, there has been a large reduction in the over intensities that also appear to be sparser besides that bottom right-side focus in comparison to their counterparts in batch one. High-intensity clusters might be related to intense solar flares or eruptions. During high solar activity periods, the likelihood of powerful solar flares or eruptions, which release substantial amounts of energy, is higher. These events result in higher-intensity readings compared to the overall spread-out intensities observed during less active periods.

3.2 Density of Hotspot Analysis

Comparing the time series plots of the hotspot discovery plots across both sets 1 and 2 with various batches. It is observed that hotspots accumulate near the edge of the Sun.

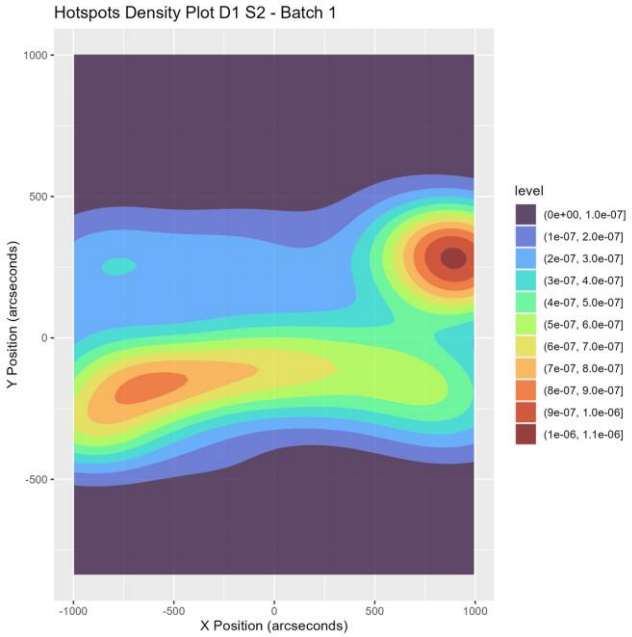

Figure 11: **Density of Hotspots, Batches 1 – 11**

In comparison to the second image (figure 12), it performs a similar pattern in batches one through eleven. Given that the plots don't show the entire Sun's surface and are just a subsection, we know that in context the hotspots are towards the Sun's equator. However, on set 2 on both d1 and d2 thresholds, we had fewer data points to map. So, batches 8 through 11 have less than 11 data points so the shape of the polygons can lead to underrepresentation or overrepresentation of specific regions, affecting the interpretation of spatial patterns.

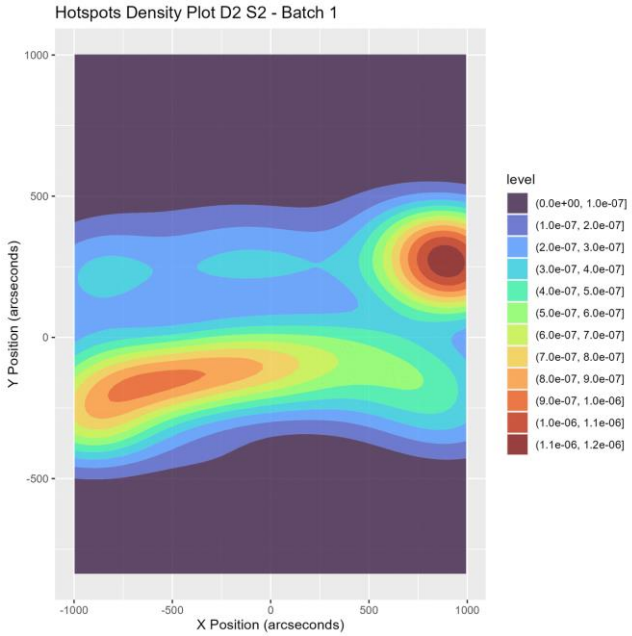

Figure 12: **Density of Hotspots, Batches 1 – 11**

3.3 Basic Statistics Comparison

The following analysis will be based on the statistics of Method 1 for Sets 1 and 2 in table format. Statistical summaries can depict essential trends in the two datasets, showcasing the dynamic variations within different batches or time periods.

In Set 1 for batches one through eleven, we see that the mean and mode behave in an exaggerated way.

Table 1: **Set 1, Method 1, Intensity Statistics**

Batch	Mean	Median	Mode	Sd of mean
1	-0.024	-0.116	0.438	-0.128
2	-0.021-0.115	-0.116	0.399	-0.122
3	-0.013	-0.114	0.504	-0.123
4	-0.007	-0.113	0.516	-0.123
5	-0.007	-0.113	0.575	-0.124
6	0.058	-0.114	2.073	-0.125
7	0.052	-0.116	2.242	-0.118
8	-0.045	-0.117	0.410	-0.117
9	-0.002	-0.114	0.759	-0.114
10	0.036	-0.111	0.880	-0.119
11	0.016	-0.111	0.668	-0.116

At the start of this cycle, we see a steady increase in mean values at small intervals for each of the batches until batch six. During this time the mode is also generally increasing which indicates a larger number of higher intensities being observed. During batches six and seven, there is a sharp increase that yields the largest mean value for all batches. During this shift, mode drastically also increases relative to the mean. Median and standard deviation appear to be unreactive to the large changes. After batch seven, normalcy returns from batches eight until eleven for both the mean and mode.

When viewing the raw statistical data for Set 2, for batches one through eleven, we see striking similarities as Set 1 with some large underlying differences from the previous set.

Table 2: **Set 2, Method 1, Intensity Statistics**

Batch	Mean	Median	Mode	Sd of mean
1	0.045	-0.175	1.326	-0.191
2	0.061	-0.163	1.360	-0.191
3	0.008	-0.162	0.867	-0.184
4	-0.036	-0.163	0.447	-0.184
5	-0.017	-0.168	0.611	-0.192
6	0.019	-0.170	0.752	-0.206
7	0.030	-0.167	0.734	-0.207
8	-0.178	-0.208	0.117	-0.214
9	-0.155	-0.205	0.156	-0.214
10	-0.152	-0.203	0.163	-0.213
11	-0.165	-0.206	0.159	-0.213

The first observation is that the mean starts positive for batches one and two and are the highest mean values for the remaining batches. This is in contrast of Set 1 where the highest mean values were not observed until batches six and seven. Starting from batch three onwards, the mean remains relatively stable until batch eight where it sharply declines at a greater scale than observed at the incline of Set 1. The median at this batch also decreases further indicating a shift in central tendency than the previous stable values. The mode for these lower batches also decreases, indicating a less amount of higher intensities in contradiction to batches one through three. Lastly, the standard deviation seems relatively unreactive to the changes by the other columns.

In comparison to Sets 1 and 2, the mean values for both sets experience peaks and lows in different batches with Set 2 experiencing a larger range of mean variation. With Set 2s larger variation of mean, median, and standard deviation it can be concluded that there was greater fluctuation or dispersion in intensity of the course of all batches in observation. This emphasizes the uncertainty and diversity in intensity measurements. When comparing the mode for both sets, Set 1 experienced an explosive increase in value not comparable in Set 2. This sharp increase indicates a larger burst of intensity over Set 2 that was not sustained.

The following analysis will be based on the statistics of Method 2 for Sets 1 and 2 in table format. Statistical summaries can depict essential trends in the two datasets, showcasing the dynamic variations within different batches or time periods.

For Table 3, in Set 1 for batches one through eleven, we see high variation of mean, mode, and standard deviation values.

Table 3: **Set 1, Method 2, Intensity Statistics**

Batch	Mean	Median	Mode	Sd of mean
1	670717	53856	3816498	10512
2	646377	53208	2649390	30360
3	815348	57936	4892813	27312
4	882901	64992	5241945	27168
5	828968	62496	5460271	23472
6	2109838	61000	40181752	18456
7	2241235	53040	44238896	42192
8	567714	46716	4521557	32664
9	1065420	59854	11315463	57168
10	1493138	73116	13383751	39312
11	1185961	72372	9266424	20016

The mean value starts low in batches one and two with the first sharp incline at batch 3, then remaining constant until batch six. During this time the median increases slightly, but the remaining attributes have large amounts of variation all resulting in higher values than in batch one. Batches 6 and 7 reflect the observations in Set 1, Method 1 where vast spikes in mean and mode are recorded in comparison to the prior

batches. This period of the cycle could be "the solar maximum, or when the Sun has the most sunsports" [1]. This is normally followed by a return to normalcy to start the cycle anew. Standard deviation follows this trend sharply increasing to its second highest value of all batches. This indicates a large burst of intensity was observed with the highest values for the data set recorded. After batch seven, we see batch eight decrease across the board with slight overall increases in batches 9 through 10 except in standard deviation.

For Table 4, in comparison Set 2 Method 1 we see similar trends for median, mode, and standard deviation.

Table 4: **Set 2, Method 2, Intensity Statistics**

Batch	Mean	Median	Mode	Sd of mean
1	694214	56508	4774875	19056
2	871203	74232	6251428	20208
3	711951	77505	5227686	41520
4	462975	79224	1608615	41400
5	566539	69875	2546199	31608
6	752380	64670	3209572	12600
7	822237	70656	3052616	11880
8	81844	10548	307899	2560
9	152760	15840	553557	2560
10	179858	17778	693665	3840
11	183831	13812	848261	3840

The mean value initially follows a similar trend for batches one and two with a slight increase followed by a decrease in batches three and four. During these first four batches, the median has increased to its highest value, indicating a shift in the central value in the positive direction. Mode slightly increases then has a dramatic decrease hinting at a lower amount of high intensity events. The standard deviation between these batches increases showing large fluctuations of intensities. From batches five to seven, the mean and mode increase significantly indicating a period of high intensity events that are greater than the prior batches. Batch 7 is the final moments of the peak intensities with batch eight having an inverse amount of activity with lowest recorded values for all attributes. From batches nine to eleven we see all attributes steadily increase except the median.

When comparing Method 2 for Sets 1 and 2, the first major difference is the range of median values. Set 1 has a range of 26,400 and Set 2 has 68,676. This indicates there is significantly more shifting on the central value for Set 2 than on Set 1.

4 Summary

For project *Helios*, summary generation was achieved successfully using Methods 1 and 2. By viewing the basic statistical attributes of the datasets, it achieved breaking down high amounts of complex data into simple variables that can be easily analyzed and reach conclusions based off of. The mapping of data is demonstrated within the density graphics and spatial variation allows us to visualize the data and make comparisons between both Set 1 and Set 2 for side-by-side analysis.

Combining all of the visualization materials and table data, the cyclic nature of the Sun can be rudimentarily interpreted and analyzed. Within the individual batches for each set, we see large amounts of building intensities that finally burst. During these largely heightened moments within the batches, the batches containing them are easily identified due to the intense amount of energy being produced that is reflected within our data. After this burst, normalcy returns back to similar levels of the prior low activity periods, restarting the cycle in preparation for another build up.

CONTRIBUTIONS

Rayyan Rahman contributed methods one and two that built the foundation of the project. Additionally, Rayyan provided contributions to sunspot recognition and generated the visual material for the final summary. Timothy Chou developed the hotspot recognition and additionally performed analysis on it. Ryan Ball performed summary generation and change analysis for the final section and report.

REFERENCES

[1] Sunspots and Solar Flares | NASA Space Place - NASA science. Retrieved November 9, 2023 from https://spaceplace.nasa.gov/solar-activity/en