

# Helios Solar Flare Discovery Report

Nikolas Velazquez  
Student ID  
University of Houston  
Houston, TX, USA  
nevelaz2@cougarnet.uh.edu

Tam Nguyen  
Student ID  
University of Houston  
Houston, TX, USA  
ttngu454@cougarnet.uh.edu

Mitchell Parks  
Student ID  
University of Houston  
Houston, TX, USA  
bmparks@cougarnet.uh.edu

Jacob Bachtarie  
Student ID  
University of Houston  
Houston, TX, USA  
jbachtar@cougarnet.uh.edu

## ABSTRACT

The Reuven Ramaty High-Energy Solar Spectroscopic Imager (RHESSI) satellite provided a broad range of data about the sun by capturing details of ray energies. This research aims to explore the data collected by this satellite, specifically the 2004-2005 and 2015-2016 subsets, and discover relevant analysis. Different methods are created in order to estimate the intensity of a solar flare given different predictor values. Using one of these methods, a hotspot discovery algorithm is created to find and track different types of hotspots, small high-intensity hotspots, and larger more regional hotspots. After comparing set 1 (2004-05) and set 2 (2015-16) datasets, it is then realized that the sunspots are having their solar cycle. Meaning the sunspot is flaring up again similar to it did 11 years ago. Researching these solar flares gives us more information to help us predict, mitigate, and understand disruptions to communication systems, satellites, power grids, etc.

## CCS CONCEPTS

- Mathematics of computing → Probability and statistics → Statistical paradigms
- Computing methodologies → Machine learning
- Information systems → Information systems applications → Data mining
- Software and its engineering → Software notations and tools → Context specific languages → Programming by example

## KEYWORDS

Reuven Ramaty High Energy Solar Spectroscopic Imager; RHESSI; sunspots; hotspots; Coronal Mass Ejections; CME; Kilo Electron Volt; keV; Kernel Density Estimation; KDE;

## ACM Reference format:

Nikolas Velazquez, Tam Nguyen, Mitchell Parks, and Jacob Bachtarie. 2023. Helios Solar Flare Discovery Report. In Proceedings of the University of Houston, College of Natural Sciences & Mathematics, Houston, TX, USA, 11 pages.

## 1. INTRODUCTION

In the vast cosmos of space, with an unprecedented number of rocky, gas, and ice planets, our galaxy, The Milky Way, is one of the many astonishing wonders of the universe. And part of that wonder is the celestial nuclear-fusion gassy star that shines close to home and in which we call, The Sun. Born within clusters of dust particles and gasses, The Solar Nebula would be the birthplace of our Sun approximately 4.6 billion years ago [2]. Composed mostly of hydrogen and helium, these clumps of dust become bigger and bigger until it collapses under its own gravity and nuclear fusion begins within the core. In ancient times, our star wasn't called The Sun, instead, ancient Greeks called the Sun *Helios* as it was widely seen as a god. It wasn't until the reign of the Roman Empire that Helios was replaced with the Latin name *Sol*, where the English word *sun* derives from [1]. Our Sun not only emits visible light, but also ultraviolet light, infrared, radio waves, X-rays, gamma rays, and even extreme electromagnetic radiation known as solar flares [3,4]. A solar flare is an intense burst of electromagnetic radiation released from magnetic energy within dark spots on the Sun known as sunspots (or hotspots), and these areas are typically cooler than other parts of the Sun's surface because it is where magnetic fields are particularly so strong that they retain some of the heat from reaching the Sun's surface [5]. The magnetic field lines near

sunspots are often tangled, crossed, and disorganized, and solar flares are created during the process of reorganization.

## 2. RHESSI & HELIOS

The RHESSI is a NASA observation satellite that was launched on February 5, 2002, and decommissioned on August 18, 2018 [6,7,8]. The RHESSI conducted several missions low orbiting around the Earth, obtaining high-resolution spectrums over a wide range of x-ray and gamma-ray energies and the ability to take images of those high energy x-rays and gamma-rays. In the 16 years of its service, it has collected many data and images of sunspots and solar flares from The Sun. The data set used in this project is a collection of solar flare information obtained between 2004-05 and 2015-16 and consists of many important attributes such as *date*, *energy.kev*, *duration*, *peak*, *total count*, *position*, etc. The '*date*' attribute is when the flare occurred located at '*x.pos.asec*' and '*y.pos.asec*', which are positions in arcsec from the center of the sun. The attribute '*radial*' is the distance from the position and the center of The Sun. The duration of each solar flare is described by the attribute '*duration.s*' which is measured in seconds from the start time '*dt.start*' and end time '*dt.end*' attributes. Amid a solar flare, there is a point in time when the solar flare emits the most energy defined by the '*dt.peak*' attribute. Other attributes such as '*peak.c/s*' measures peak count rate in correct counts (peak counts/seconds), '*total.counts*' is the total counts in corrected counts of the number of energy range 6-12 keV integrated throughout flare summed over all sub collimators (including background), and '*energy.kev*' (subdivided into lower bound '*energy.kev.i*' and upper bound '*energy.kev.f*') is the highest energy range in which the flare was observed measured in kilo electron volt (keV).

### 3.1 METHODOLOGIES

Our interpretation of how to develop two methods of solar flare intensity estimation based on a set of flare events were to build two regression models. Method 1's predictor variables: '*duration.s*', '*peak.c/s*', '*x.pos.asec*', '*y.pos.asec*', '*radial*', '*energy.kev.i*', '*energy.kev.f*', '*day*', '*month*', '*year*', '*active.region.ar*' with the purpose of predicting the target variable,

'*total.counts*'. Method 2's predictor features: '*day*', '*month*', '*year*', '*active.region.ar*', '*peak.c/s*', '*total.counts*', '*x.pos.asec*', '*y.pos.asec*', '*radial*' to predict the '*intensity*' attribute which is composed of the average of an observations '*energy.kev*' value multiplied by the '*duration.s*' attribute. The snippet below shows the formula for the target variable of Method 2.

```
def addIntensity(data):
    avg_energy = (data['energy.kev.i'] + data['energy.kev.f'])/2
    data['intensity'] = avg_energy * data['duration.s']
    return data
```

**Figure 1: function for creating intensity target parameter for method 2**

We started by trying models we already knew and had discussed in class: support vector machines, k-nearest neighbors, linear regression, and decision trees. From the *Scikit-learn* library in *Python*, we utilized the *GridSearchCV* searching algorithm initially to find the optimal parameters for the aforementioned models [9]. Our results showed that these models were failing to capture meaningful patterns; additionally, models like *SVR()* and *DecisionTreeRegressor()* were very computationally expensive when fitting to the data.

These factors led us to change our approach; we pivoted towards learning about different models, how to validate regression models, types of hyperparameter tuning algorithms and using data preprocessing tools. The next set of models we tested were gradient boosting algorithms like *GradientBoostingRegressor()* from *sklearn.ensemble* and *XGBRegressor()* from the *xgboost* library, *RandomForestRegressor()*, and different variations of linear regression such as: polynomial regression, ridge regression, and lasso regression. We opted to modify the scoring hyperparameter within our searching algorithm from the default *r2\_score* to *neg\_mean\_absolute\_error*. This decision stems from our objective of identifying a model that excels in accurately predicting the target variable. We believe that *neg\_mean\_absolute\_error* aligns more closely with our goal, emphasizing the importance of minimizing absolute prediction errors and enhancing the overall predictive performance of our models. We improved the cross-validation strategy in our tuning algorithm by incorporating *RepeatedKFold()*, a method that repeats the k-fold cross-validation

process multiple times. This not only provides a more comprehensive evaluation but also contributes to the robustness of our results. By setting a random seed within our validation function, we are ensuring reproducibility across different runs, making our results more reliable and consistent with each function call.

```
def cv(model):
    cv = RepeatedKFold(n_splits=10, n_repeats=3, random_state=1)
    # evaluate model
    scores = cross_val_score(model, X_train, y_train,
                             scoring='neg_mean_absolute_error', cv=cv, n_jobs=-1)

    #for cross_val being used:
    scores = np.absolute(scores)
    print(scores)
    print('Mean MAE: %.3f (%.3f)' % (scores.mean(), scores.std()))
```

**Figure 2:** function used to cross-validate tuned models

After researching alternative searching algorithms, we decided to change the hyperparameter tuning algorithm from *GridSearchCV* to *RandomizedSearchCV* because of its faster runtime without compromising performance. Unlike the exhaustive search approach of *GridSearchCV*, *RandomizedSearchCV* employs a randomized sampling strategy, allowing us to explore the hyperparameter space more efficiently [9]. One key feature contributing to its efficiency is the ability to take probability distributions as the parameter list. This flexibility in specifying parameter distributions enables a more data-driven exploration, potentially leading to the discovery of optimal hyperparameter configurations that might not have been considered in the exhaustive grid search conducted by *GridSearchCV*.

```
xg_params = {
    'n_estimators' : randint(100,1201),
    'max_depth' : randint(1,11),
    'learning_rate' : uniform(0.001, 0.3)
}
```

**Figure 3:** Example of the param\_list for the *XGBRegressor()* model

Noticing that some models tested are sensitive to the magnitude of features like those of *SVR()* and variations of *LinearRegression()*, we implemented preprocessing tools like *RobustScaler()* and *StandardScaler()* to test if the model's performance would improve. Unexpectedly, the models did not show any improvement and either resulted in the

same if not a lower score. Evidently, because the models with scaled data and unscaled data resulted in the same scores we can assume *Scikit-learn's* implementation includes an internal normalization step during the training process.

### 3.2 TASK 1 DISCUSSION

Our results showed that the best-performing model for method 1 was the *XGBRegressor()* with the parameters: learning\_rate= 0.248, max\_depth= 10, n\_estimators= 546; the parameters were found from the result of *figure 3*. The *XGBRegressor()* model outperformed all other models by a fair margin except for *RandomForestRegressor()* which performed similarly.

The intensity maps are developed as 4 types of files:

#### 1: Intensity Map png

Folder: Intensity\_Map\_PNGs

File\_Name:intensity\_map\_Method

N\_BatchN\_startyear\_endyear.png

Description: Scatter plot of intensity map according to method number, batch number, startyear and endyear. Plots all the points for the year.

#### 2: Timeline Gifs of the Intensity Maps

Folder: GIFs

File\_Name: Method N BatchN startyear\_endyear.gif

#### 3: Timeline MP4s of the Intensity Maps

Folder: MP4s

File\_Name: MethodN BatchN startyear\_endyear.mp4

#### 4: PNGs used for GIFs and MP4

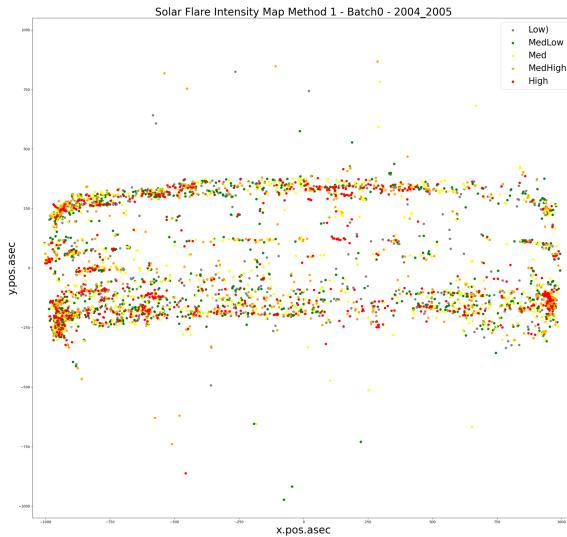
Folder: Method N BatchN startyear\_endyear pngs

File\_Name: YYYY-MM-DD

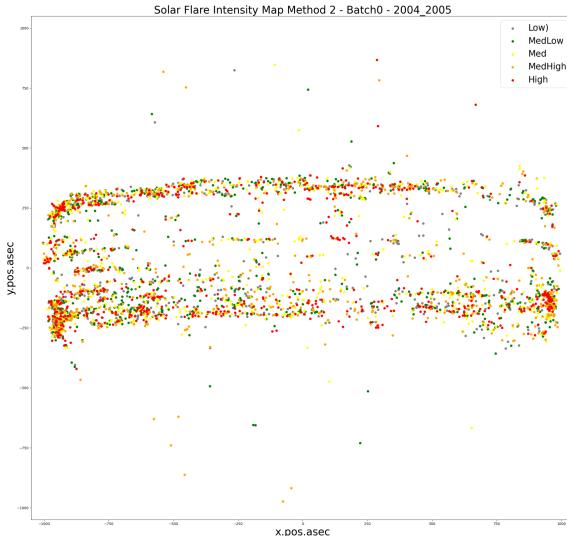
Note: All files have a label top right of the plot that indicates the intensity category of the points. PNG folders can be used to reference specific dates as needed. **Both MP4s and GIFs are intensity timelines** for the year. Displaying the plots on a black circle that is considered the sun. The files are stored in their respective folders, the folders are auto-created upon creation. We have also included the naming convention of the files. "N" would be the method or batch number, while startyear and endyear are in the YYYY format with the startyear and endyear being only 1 year apart. (2004-2005 or 2015-2016)

Intensity maps for **months 1+2+3+4 are considered as Batch0; months 21+22+23+24 are considered as Batch10**. For all 4 types of files, they are created for each batch and each method we are comparing.

The full-size intensity plot is in the respective folder.



**Figure 4: intensity map Method 1 Batch 0 2004-2005**



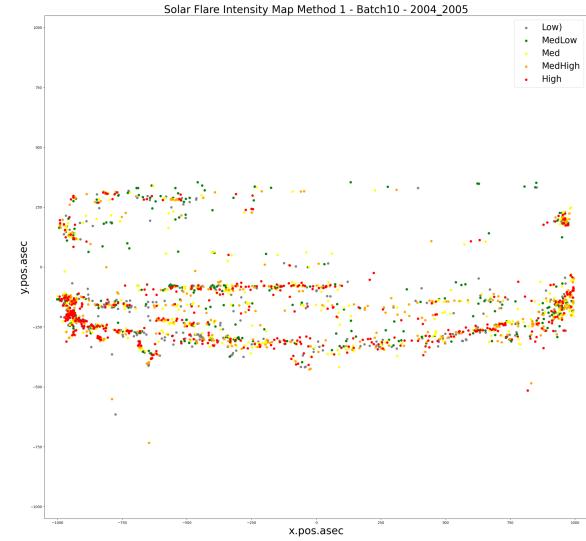
**Figure 5: intensity map Method 2 Batch 0 2004-2005. Very similar to Method 1 Batch 0 2004-2005.**

8 intensity maps were created. Batch0 for methods 1 and 2. Batch10 for methods 1 and 2.

For both Batch 0 and Batch 10 for their respective year and methods, there does not seem to be any significant differences aside from the number of intensity categories. ie: More Low-High categories. Example: Batch 0 2004-2005 method 1 vs method 2, has no significant differences.

This is because the models are different, resulting in different measures or predictions for the intensities of

the plots. However, if we compare Batch 0 to Batch 10 for any of the years, there are significant differences. Batch 0 for 2004-2005, points are generally uniform along the center y, across the x-axis. Batch 10 for the same year is like Batch 0, but there are fewer points in the upper right quadrant of the plot. This suggests that Batch 0 is more diverse on the sun in solar events than Batch 10. This will be further explained after the table.



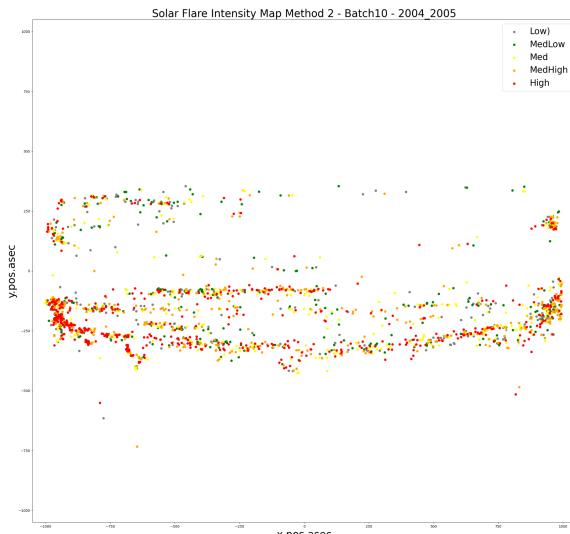
**Figure 6: intensity map Method 1: Batch 10 2004-2005**

Categories	Batch 0	Batch 10
Low	713	355
MedLow	828	337
Med	776	394
MedHigh	658	389
High	687	489
Total	3662	1964

**Table1 for Method 1 2004-2005**

Categories	Percentiles
Low	Intensity $\leq 20^{\text{th}\%}$
MedLow	$20^{\text{th}\%} < \text{intensity} \leq 40^{\text{th}\%}$
Med	$40^{\text{th}\%} < \text{intensity} \leq 60^{\text{th}\%}$
MedHigh	$60^{\text{th}\%} < \text{intensity} \leq 80^{\text{th}\%}$
High	$80^{\text{th}\%} < \text{intensity}$

**Table 2 for Intensity Categories/Partitions**



**Figure 7: intensity map Method 2: Batch 10 2004-2005. Similar to Method 1 Batch 10 2004-2005.**

Batch 0 has more intensity points and higher categorical values than Batch 10. This suggests that there was more solar activity during the period of Batch 0. There are also more overall medlow-high categories, which also suggests more occurrences of solar flares with greater intensities than that of Batch 10. However, Batch 10's High category is 33.67% different from Batch 0. This suggests that Batch 10 had some intense moments, but not as frequent as Batch 0.

### 3.3 TASK 2 DISCUSSION

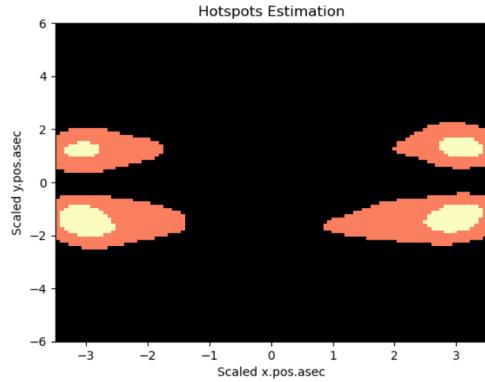
We approached the topic of creating a hotspot discovery algorithm by the means of using a Kernel Density Estimation. First, the data is partitioned by the intensity estimation developed in method 1. These

partitions are split by the differing levels of intensity, in order to develop hotspots where more intense solar flares are more common. These partitions were of low, medlow, med, medhigh, and high with each containing 20% of the overall data set.

From the `scipy.stats` library in Python, we used the `gaussian_kde` function in order to calculate a kernel density estimation(KDE) of the model. Once we had a density function, we wanted to find a way to visualize the density plot. To do this, we created a 100x100 grid with the same bounds of the data set for `x.pos.asec` and `y.pos.asec` attributes. The bounds of the space are (-1007, 1005) for `x.pos.asec` and (-998, 1012) for `y.pos.asec`. Once the 100x100 mesh grid is created, we use the kernel density estimation function in order to predict the average value for each of the squares within the grid. Each grid's density is calculated to be the average of the kernel density function over the 2d interval.

This allowed us to analyze the density of a finite amount of points on the graph in order to determine values for two different hotspot thresholds: D1 and D2. D1 to define a density of very small hot spots with D2 to be a density threshold of larger regional hotspots. We chose D1 to be considered the previously mentioned high partition of the dataset and the D2 boundary to be the medium-high and high parts of the dataset. D1 was chosen to be 241486 and D2 was chosen to be 82637.

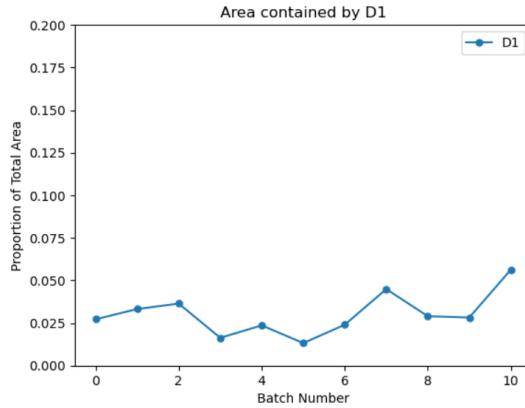
The KDE for both of these is used to find the areas the function/s considers hotspots. The plot is then made into a graph with the two different types of hotspots to be determined. The yellow areas represent very small hot spots meeting D1 whereas the orange areas are areas that do not hit threshold D1 but do hit threshold D2. Both the X and Y coordinates are shown in their scaled/z-score state. The graph shown is the visualization of the density estimation for the entirety of the 2004-2005 data set.



**Figure 8: Time series plot of the area confined by D1 for each of the 11 batches.**

As can be seen in the above visualization, there are 4 major concentrations of hotspots. These areas look similar to the above intensity plots in which most of the solar flares are within a belt around the center of the y.pos.asec variable. It can be seen that the hotspots found are primarily around the center of the y.pos.asec, however, are on the far ends of both sides of the x.pos.asec variable.

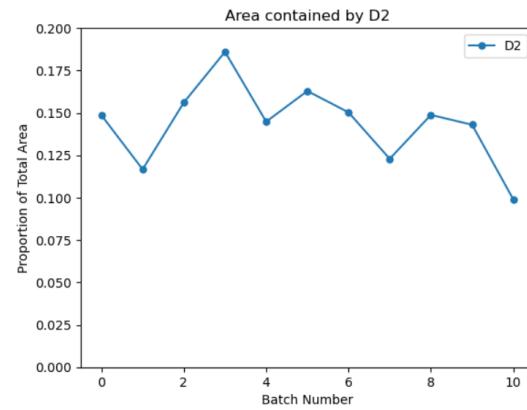
For each of the 11 batches mentioned above, a KDE model was made. While making the model, a counter kept track of the number of grid spaces from the 100x100 mesh grid that were greater than D1(yellow area) as well as the area contained between D1 and D2(orange area).



**Figure 9: Time series plot of the area confined by D1 for each of the 11 batches.**

The amount of grids where the grid's average value was higher than the D1 is represented in the time series above. The x-axis represents the batch number and the y-axis represents the proportion of the total proportion of area that is represented by D1. Batch 10

and batch 7 have significantly more area than the other batches. Interestingly, batch 4 which is the months of September through December of 2004 is on the lower end of the area contained by hotspots, while being the same time of year as the highest, batch 10, just a year earlier. As the sun does not revolve (with respect to Earth) and rotates around its axis around every 27 days so there being little-no correlation between years is understandable. The lowest area contained is batches 3 and 5 representing the may-august of 2004 and september-december 2004.

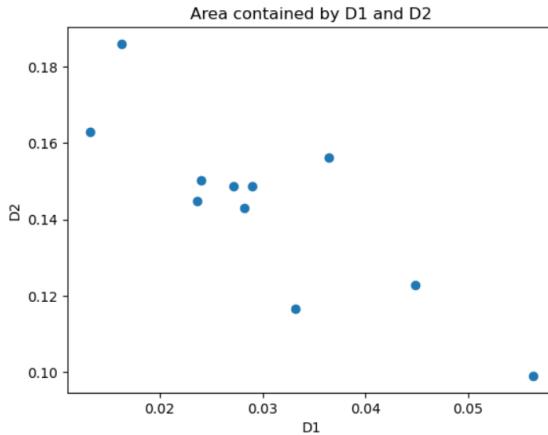


**Figure 10: Time series plot of the area confined by D2 for each of the 11 batches**

Because D2 has larger, more regional hotspots, the areas are significantly larger in proportion to the total area compared to the area confined by D1. The highest months are batch 3 and batch 5 with the lowest being batches 1 and 10. Similarly to the previous time series, there is little correlation between batches a year away from each other.

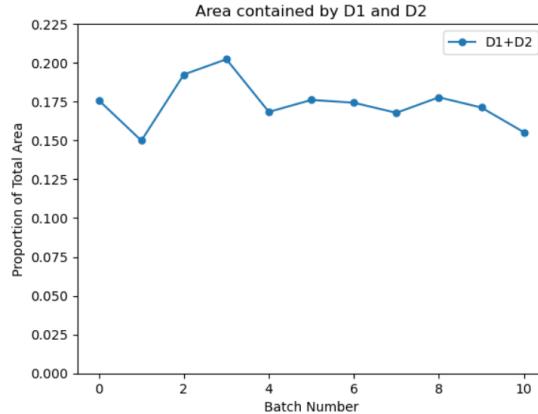
From the analysis of these two time series, we realized that the highest months for the area contained by D1 was the lowest contained by D2 and similar for the lower of D1 and the higher of D2. This encouraged us to develop a plot of the two areas added together as their relationship seems to be inversely proportional, that is the more area confined by D1 threshold, the less area that is confined by D2

threshold.



**Figure 11: Scatterplot of total area proportions for D1 and D2**

The Table supports our theory about the area within D1 and the area between D1 and D2 being negatively correlated. The correlation coefficient of D1 and D2 is -0.84 with a coefficient of determination of 0.71. The two areas have a strong negative correlation between the two.



**Figure 12: Time series plot of the area confined by D1+D2 for each of the 11 batches**

The Table above shows the area confined by the sum of the D1 and D2 areas together. The time series is much more stagnant starting in batch 4 onwards as they all have a density around 0.15-0.175. The first 4 batches(0 through 3) are much more variable. Batch 1 is much lower in proportional area than the area around it. Alternatively, batches 3 and 4 have a much higher proportional area of around 0.2 than all the rest of the batches.

### 3.4 TASK 3 DISCUSSION

When comparing the two figures, there is a similarity of clusters between the two dates. The plots do not fall on the same dates, because the Sun rotates on its axis every 27 days. So there would be a need to calculate the difference in days, to estimate the general dates to compare. For Figure 13 2004-01-06, and Figure 14 2015-01-03, the difference in days is 4015 calendar days.

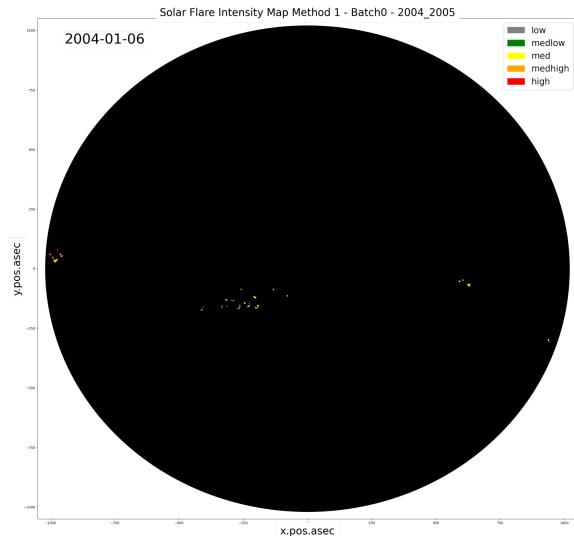
$$\text{Rotations} = 4015/27 = 148.704 \text{ rotations}$$

$$149 \text{ rotations} = 4023 \text{ days.}$$

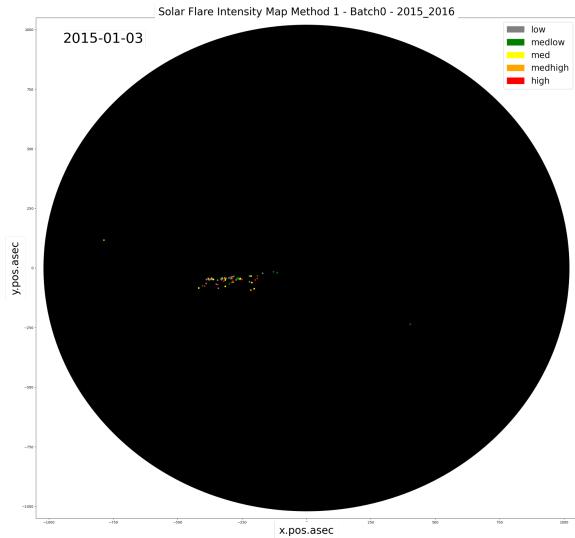
$$\text{Difference of days} = 4023 - 4015 = 8 \text{ days}$$

Getting an 8-day difference. We can give or take some days for variation. In our case of Figures 13 and 14, we will take a 3-day difference.

In Figures 13 and 14, it is correct to say they are of the same solar sunspots. However, there are differences in the location of the sunspot or intensity of the sunspots. The reason why is that it is the same sunspot is because “the number of sunspots has long been known to vary with an approximately 11-year repetition known as solar cycle”.[13] If we use the gifs or mp4 files to look at the timeline from 2004-01-06 to 2004-01-15, and 2015-01-03 to 2015-01-12. We can see the same sunspots on the sun from the solar cycle.



**Figure 13: Method 1, Batch 0, 2004-01-06**  
**GIF 1/MP4 1 in appendix A.3.4.**



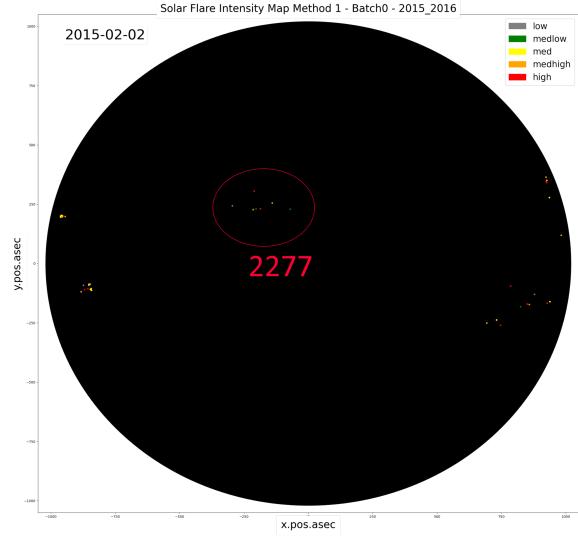
**Figure 14: Method 1, Batch 0, 2015-01-03  
GIF 2/MP4 2 in appendix A.3.4.**

Computing basic statistics for set 1 and set 2 we can observe that the number of observations within sets 1 & 2 is 17,506 and 10,779; Their respective intensity means are 7,418.51 and 9,760.98 . The lower mean of set 1 in comparison to set 2 can be explained by the higher count in set 1.

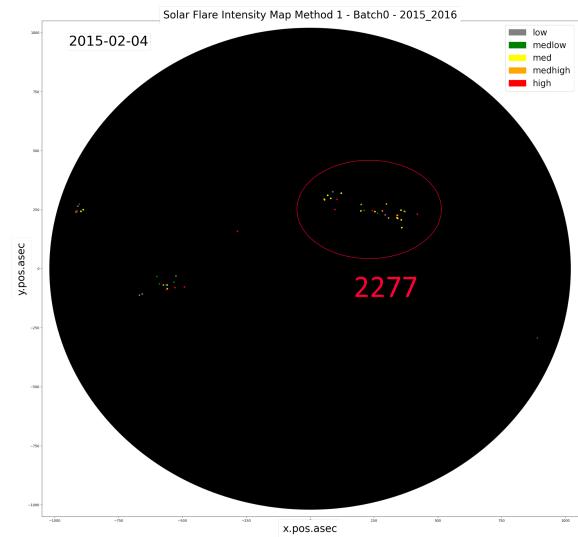
Dates range 2004-01-06 to 2004-01-15, has 574 observations with the average of the predicted ‘total.counts’ attribute being 288,100.27. Dates range 2015-01-03 to 2015-01-12 has 476 observations with the average of the predicted ‘total.counts’ being 171,861.87. This suggests that the year 2004 was more intense and had more solar activity than the 2015 solar cycle.

## 4 DATA DISCOVERY EXPLORATION DISCUSSION

In any of the png, gif, or mp4 files, there will be clusters of intensity points throughout the timeline. These clusters are sunspots from the sun, and some of them are solar flares. The clusters also move over time across the sun. This happens because the sun is rotating, and the sunspots are moving along with the sun’s rotation. Some clusters would have an increase in points or intensity, signifying solar activity or solar flares. A depiction of such a case is shown in the year 2015 from February-02 to February-04 in figures 15 and 16. In this specific case, February 02, has 7 intensity points near the center of the sun. Feb 04, has 19 intensity points near the center of the sun but shifted right.



**Figure 15: Sunspot 2277 Method 1, Batch 0, 2015-02-02.**

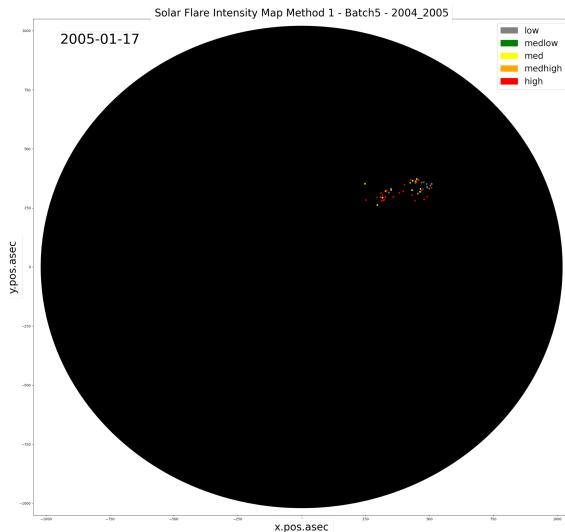


**Figure 16: Method 1, Batch 0, 2015-02-04.  
Clusters have shifted as the sun rotates. There are more intensity points as well.**

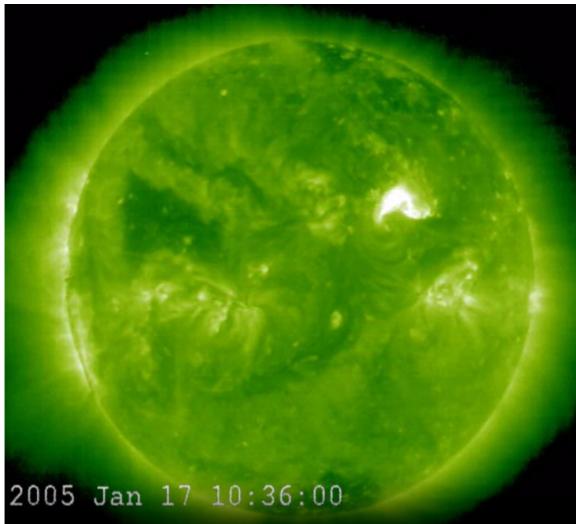
The sharp difference in intensity point tells us that region 2277 is being active. It constitutes a localized solar flare on the sun. This however does not constitute a Coronal Mass Ejection. One example of CME is mentioned in January 2005 [10].

“In January 2005, the most intense solar flare in 15 years with sunspot 720 erupting, 5 times from 15th to 20th” [10].

Using method 1, we find that this date aligns with batch 5 of 2004-2005. Figure 17 displays the same date with the intensity points on the sun.. This aligns with the image from NASA's image of Figure 18[11]. There are 55 intensity points on Jan 17th. Jan 20th has 53 intensity points. This tells us **that more intensity points do not calculate the highest intensity values**. However, it can be considered that a solar storm or some solar activity is occurring. In this case, this is a solar flare that is accompanied by a CME [12]. On Jan 20th the highest intensity in 15 years was recorded as 318,884,832 total.counts. Our method 1 model predicted 318,884,567 total.counts. This is an 8.31% difference. Our model performed as expected. In Figure 20, the high amounts of static depicts the CME hitting the satellite.

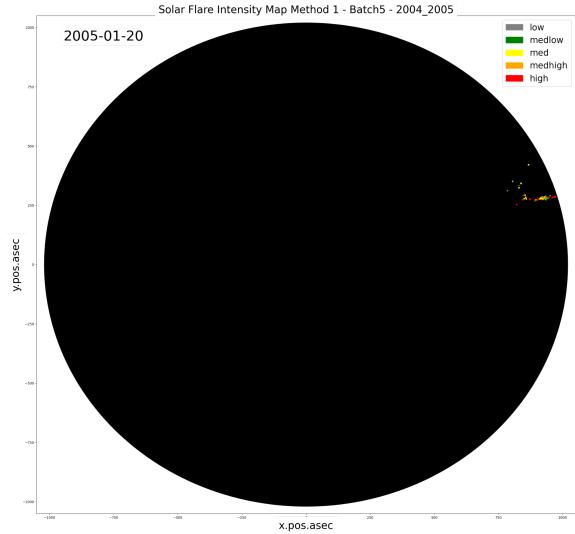


**Figure 17: Method 1, Batch 5, 2005-01-17**

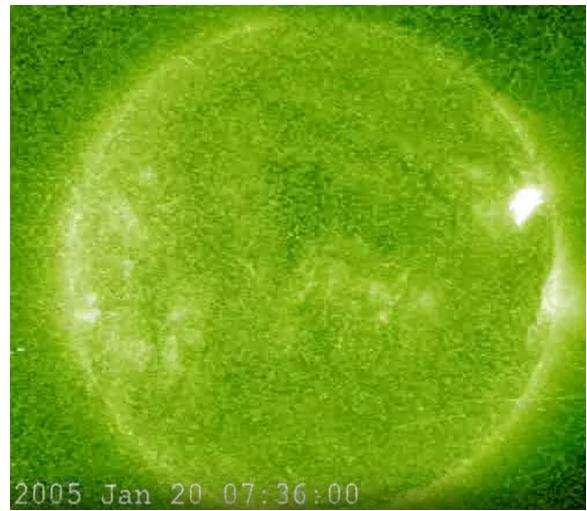


**Figure 18:** The sunspot region 720 is

**currently exploding on Jan 17 taken from SOHO/EIT Satellite.<sup>1</sup>**



**Figure 19: Method 1, Batch5, 2005-01-20**



**Figure 20: Highest intensity sunspot region 720 exploding on Jan 20.<sup>2</sup>**

---

<sup>1</sup> Bridgman, T. (2005) *January 2005 Solar Flares from SOHO/EIT, NASA*. Available at: <https://svs.gsfc.nasa.gov/3160>

<sup>2</sup> Bridgman, T. (2005) *January 2005 Solar Flares from SOHO/EIT, NASA*. Available at: <https://svs.gsfc.nasa.gov/3160>

## 5. REFERENCES

- [1] Owen, D. (no date) *How did the Sun get its name?* Available at: <https://www.spacecentre.nz/resources/faq/solar-system/sun/name.html>. (Accessed: 11 November 2023).
- [2] *What is a nebula?* (2022) NASA. Available at: <https://spaceplace.nasa.gov/nebula/en/> (Accessed: 11 November 2023).
- [3] *What are solar flares?* (no date) ESA. Available at: [https://www.esa.int/Science\\_Exploration/Space\\_Science/What\\_are\\_solar\\_flares](https://www.esa.int/Science_Exploration/Space_Science/What_are_solar_flares) (Accessed: 11 November 2023).
- [4] *What is a solar flare?* (no date) NASA. Available at: <https://www.nasa.gov/image-article/what-solar-flare/> (Accessed: 11 November 2023).
- [5] *Sunspots and solar flares* (2021) NASA. Available at: <https://spaceplace.nasa.gov/solar-activity/en/> (Accessed: 11 November 2023).
- [6] *RHESSI: A High-Resolution Spectroscopic Imager* (no date) NASA. Available at: <https://hesperia.gsfc.nasa.gov/sfttheory/imager.htm> (Accessed: 11 November 2023).
- [7] Meaney, C. (2008) *RHESSI Spacecraft*, NASA. Available at: <https://svs.gsfc.nasa.gov/20163/> (Accessed: 11 November 2023).
- [8] *RHESSI RETIRES! RHESSI'S Greatest Hits 2002 - 2018* (no date) NASA. Available at: <https://hesperia.gsfc.nasa.gov/rhessi3/> (Accessed: 11 November 2023).
- [9] Pedregosa, F. et al. (2011) *Scikit-learn: Machine Learning in Python*, *Journal of Machine Learning Research*. Available at: <https://scikit-learn.org/stable/about.html#citing-scikit-learn> (Accessed: 11 November 2023).
- [10] List of solar storms (2023) Wikipedia. Available at: [https://en.m.wikipedia.org/wiki/List\\_of\\_solar\\_storms](https://en.m.wikipedia.org/wiki/List_of_solar_storms) (Accessed: 11 November 2023).
- [11] Bridgman, T. (2005) *January 2005 Solar Flares from SOHO/EIT*, NASA. Available at: <https://svs.gsfc.nasa.gov/3160> (Accessed: 11 November 2023).
- [12] McKenna-Lawlor, S. et al. (2010) *Moderate geomagnetic storm (21–22 January 2005) triggered*

*by an outstanding coronal mass ejection viewed via energetic neutral atoms*, *Advancing Earth and Space Sciences* . Available at: <https://doi.org/10.1029/2009JA014663> (Accessed: 11 November 2023).

[13] *Sunspots/solar cycle* (no date) *Sunspots/Solar Cycle | NOAA / NWS Space Weather Prediction Center*. Available at: <https://www.swpc.noaa.gov/phenomena/sunspotssolar-cycle> (Accessed: 11 November 2023).

## APPENDIX

### A. FILE PATH

#### A.1 Introduction

#### A.2 RHESSI & Helios

#### A.3.1 Methodologies

#### A.3.2 Task 1 Discussion

#### A.3.3 Task 2 Discussion

#### A.3.4 Task 3 Discussion

GIF 1: Method 1 Batch0 2004\_2005.gif

GIF 2: Method 1 Batch0 2015\_2016.gif

MP4 1: Method 1 Batch0 2004\_2005.mp4

MP4 2: Method 1 Batch0 2015\_2016.mp4

### A.4 Data Discovery Exploration Discussion

### A.5 References

## CONTRIBUTIONS

Nikolas Velazquez — Formatted document, helped develop task 2 algorithm, found helpful programming references, wrote Introduction, RHESSI & Helios, and helped write Task 3 Discussion.

Mitchell Parks — Developed algorithm and wrote discussion for task 2 as well as intensity categories/partitions. Developed abstract.

Jacob Bachtarie — Developed intensity estimation methods 1 & 2, parameter tuning algorithm, cross-validation algorithm, model discovery, wrote

methodologies and helped write task 1 & 3 discussion.

Tam Nguyen — Task 1 part 3-5 and Task 3 discussion, Data Discovery Exploration Discussion, Animations, PNGs creation and Batch creations.