York University
PHYS3130 W2025

Analysis Of The Effects Of Solar Flares On Satellites And GPS

Matteo Migliorati
`Final Code .ipynb: Final_Code.ipynb`
url: `https://github.com/MatteoM04/Phys3130_MM`
April 16 2025

1 Introduction

Solar flares are some of the most powerful and yet potentially disruptive natural phenomena in our solar system. These sudden bursts of electromagnetic radiation produced by the sun are composed of x-rays and high energy particles. Solar flares occur when there are magnetic reconnection events that happen on the sun's surface (Gopalswamy, 2010). When the flares erupt from the sun, the energy directed towards earth can significantly impact the magnetosphere and ionosphere. This interaction that it has with earth can have many profound consequences for humanity. The interaction directly affects a wide range of technology here on earth and in orbit. However, the most significant technology that solar flares affect is our satellite systems and GPS navigation. (Khabarova, 2024) The satellites that are created are designed to operate only within specific tolerance thresholds in relation to radiation. Majority of the commercial and scientific satellites that are sent into earth's orbit can only withstand approximately up to 8000 electron volts (eV) (Khodairy et al., 2020). Satellites that are exposed to anything substantial over 8000 electron volts can result in permanent hardware damage (Khodairy et al., 2020). Having permanent hardware damage can lead to large errors in data being collected and ultimately having to retire the satellite. Unfortunately, GPS systems will have a similar fate. Having increased ionization in the ionosphere caused by the solar flare will interrupt GPS information that passes through earth's ionosphere (Wu, 2016). This results in errors in the accuracy of the GPS system and in severe cases, can lead to complete distortion and loss of the signal (Shigeoka et al., 2004). Having these disruptions directly affect many different industries such as the aviation industry, military, and emergency services (Wu, 2016). The purpose of this study is to identify the uniquely high energy level events that occurred during the March 2015 solar storm. This will be done by using unsupervised machine learning techniques. The isolation of outliers in the data can be done to compare them with normal amounts of solar radiation.

2 Methodology

The March 2015 solar storm was one of the most intense solar weather events ever recorded (Ray et al., 2016). During this solar storm, some of the flares and coronal mass ejections from the sun were so extreme that they reached the highest class of solar flare (X-class) (Ray et al., 2016). Although this was very dangerous for earth's magnetic field, it also created a terrific opportunity to analyze solar flare data from the sun (Ray et al., 2016).

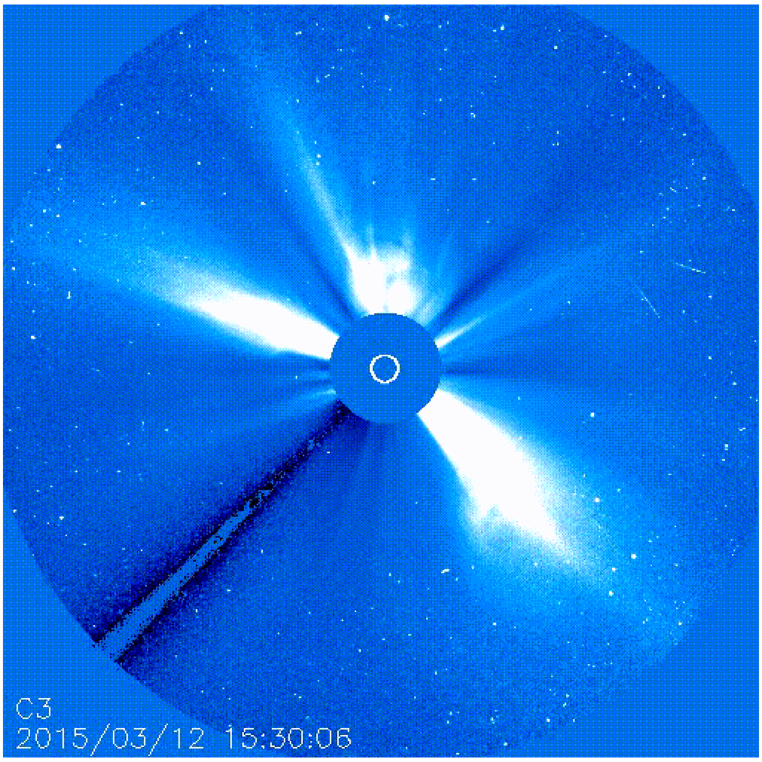

Figure 1: Picture of X-class Flare during March 2015 Solar Storm (Virtual Solar Observatory, 2025)

The dataset that is used in this research was from the March 2015 solar storm, specifically the 17th and 18th during solar cycle 24 (Ray et al., 2016), and was obtained from the Virtual Solar Observatory (Virtual Solar Observatory, 2025). The Virtual Solar observatory displays an extensive archive of solar flare data that is collected from many different instruments (Virtual Solar Observatory, 2025). There is a large variety of some instruments that are based here on earth and some that are based in earth's orbit. The main instruments that collected the data are RHESSI, SOHO, STEREO A/B and SDO (Virtual Solar Observatory, 2025). The data employs quantitative characteristics and aligns with a positivist research paradigm. The dataset includes relevant details such as the instrument used to collect the data, the maximum and minimum spectral energy in electron volts and time the flare starts and ends. The dataset comprises of around 14400 rows which makes it very suitable for data analysis and machine learning purposes. The large volume of data allows the use of machine learning to be applied with confidence. With such a big amount of data, there were some needs for data cleaning. The data cleaning was done in python and the cleaning that was done was to eliminate missing values, convert all energy values into electron volts and convert the timestamps into the standard format. With the data cleaned, it is then possible to use machine learning for the dataset. The core of the machine learning technique that is used in this study is called the local outlier factor (LOF) technique. The local outlier factor is an unsupervised outlier detector method that compares the local density of a data point and compares it to the densities of its neighboring data points (Pedregosa et al., 2011). If there is a data point that is considered to be in a lower density region compared to its neighboring data points, it will be flagged as an outlier. In the code used, the nearest neighbor's number was 20. That value gave a good balance between being sensitive enough to identify the extreme points but not being too sensitive to misidentify others that should not be considered anomalies (Pedregosa et al., 2011). This method was chosen because of its effectiveness in identifying unique data points in the dataset. Although this research has a focus on the maximum spectral energy emitted by the sun, some future work that can be done could include using other magnetic field data to identify how much energy earth deflected away.

3 Exploratory Data Analysis and Results

The exploratory data analysis (EDA) phase of the data that was collected has revealed many important patterns and characteristics that can be viewed. The primary variable that was analyzed was the maximum spectral energy (in electron volts) as it shows the most intense energy of the solar flares that occurred during the 17th to the 18th of March 2015.

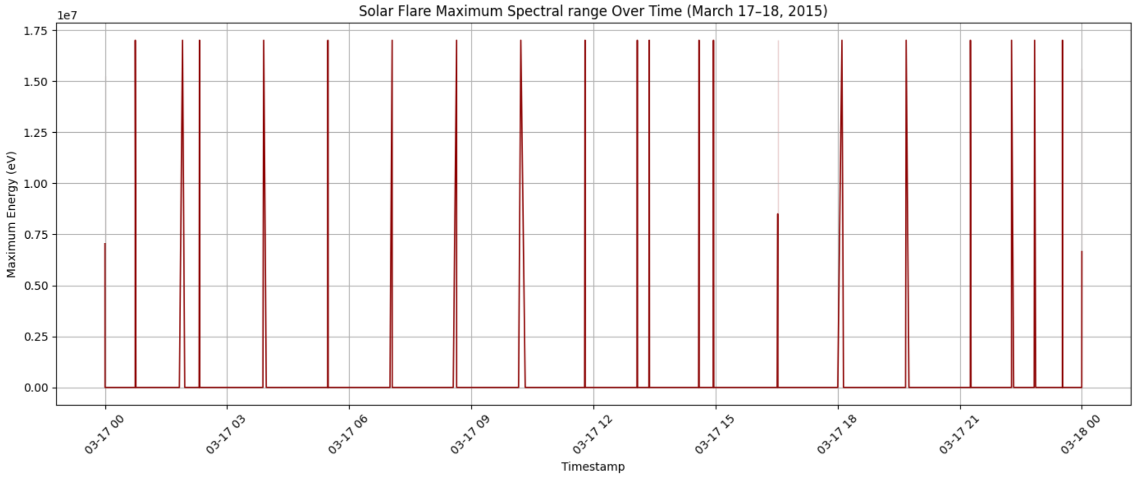

Figure 2: Initial Data Visualization of Max Spectral Energy vs Time

The initial look at the dataset showed that energy levels fluctuated hastily. The data showed values that reached as low as a couple of electron volts up to 100 million electron volts. This large range of data required data cleaning as mentioned above (eliminating missing values, conversion of all energy values into electron volts and conversion of the timestamps into the standard format). After the data cleaning, the descriptive statistics have shown that the mean of the maximum spectral energy is approximately 130,000 electron volts, a median of 3.12 electron volts and a standard deviation value of 2,400,000 electron volts. Having a small median compared to the larger mean confirms the expectation of have some largely skewed points in the data.

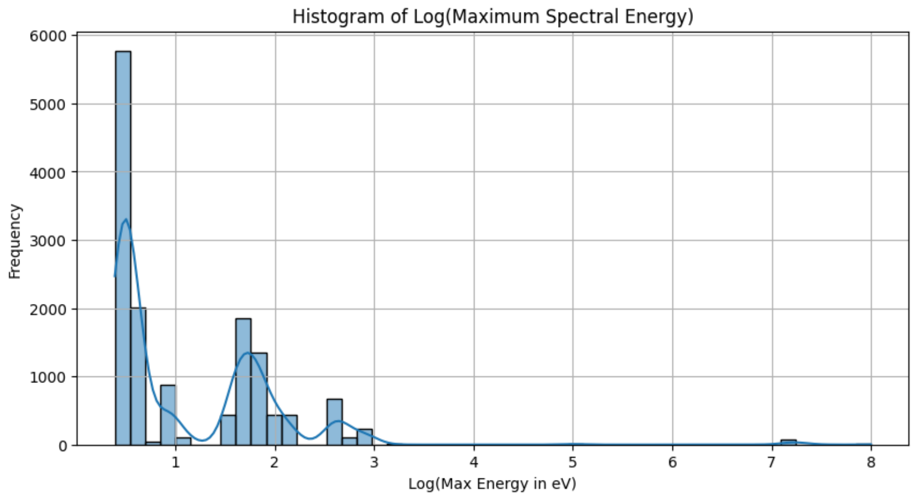

Figure 3: Log Transformed Version of Frequency and Maximum Spectral Range

A log transformed histogram of max spectral energy vs frequency was created to help further visualize the skewness of the data. The log transformed histogram was used instead of a regular one to effectively compress the data and make it easier to see certain patterns. The figure above displays a denser cluster of energy between the 1 and 100 electron volt range, however, having a scarce point on the 100 million electron volt frequency range. This is significant because it indicates the presence of the rare X-class flares (Ray et al., 2016). These observations of the X-class flares are the main concern as they are the flares that will deeply affect and exceed the maximum tolerances of the satellite and GPS systems the most. The histogram also shows that the data does not follow a normal distribution due to its randomness.

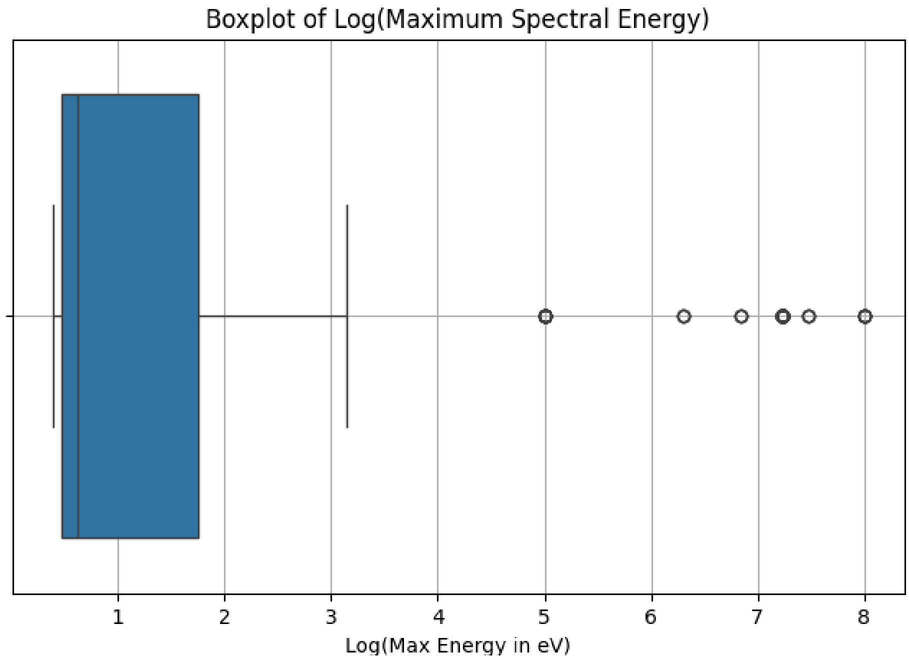

Figure 4: Log Transformed Version of Box plot for Maximum Spectral Range

The log transformed boxplot in the figure above using the maximum spectral energy helps to visualize the spread of data and detect any extreme outliers. Majority of the points are within the interquartile range, however, some points also are displayed past the upper whisker of the boxplot. These points are considered the extreme outliers in the data.

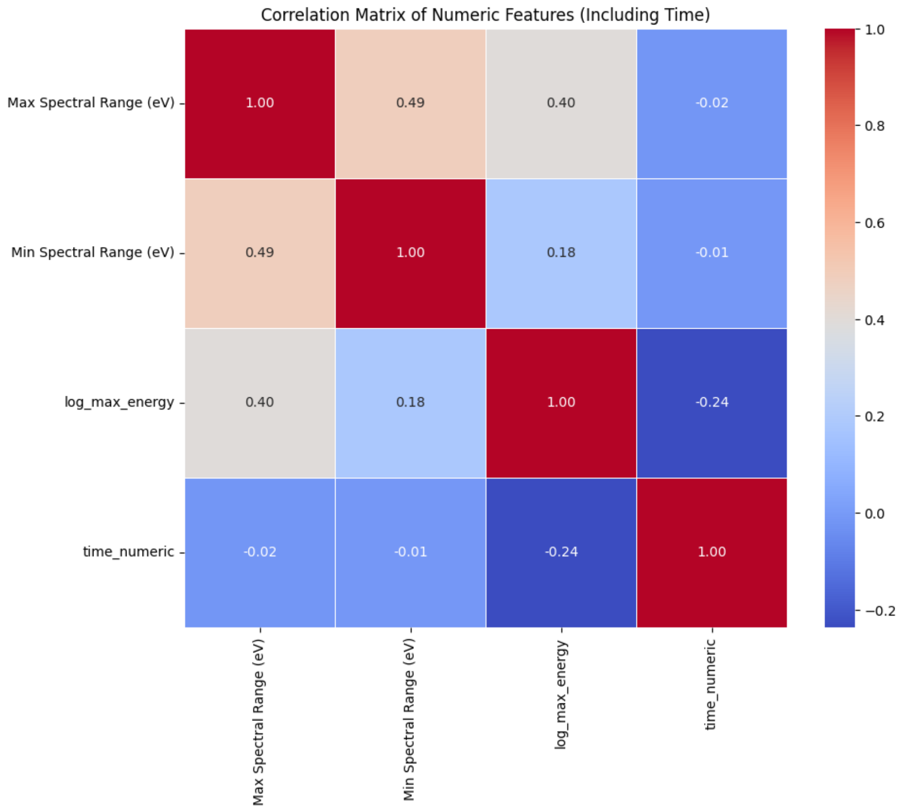

Figure 5: Correlation Matrix Comparing Variables

The correlation matrix was generated to compare the log and normal maximum spectral energy to other features in the dataset. The correlation matrix shows that there is a weak negative correlation between time and the log of the max spectral energy. This could imply that the flare energy could be decreasing over time. The correlation matrix also shows that there is some correlation between maximum and minimum spectral range. This can imply that when the maximum spectral energy is high, the minimum spectral range would also be high.

Having the extremely skewed data of the unique X-class solar flares and weak correlations make this data suitable for unsupervised learning for anomaly detection (Pedregosa et al., 2011). Supervised learning models would be less suitable for this data. Therefore, the decision to use the local outlier factor machine learning model is theoretically and practically more justified. After the processing and conducting of an exploratory data analysis, the next step is to involve applying the local outlier factor algorithm. The local outlier factor is a density based and unsupervised detection method that will calculate the local density of a data point and compare it to the density of its neighboring points (Pedregosa et al., 2011). Data points that are in lower density regions will be flagged as outliers.

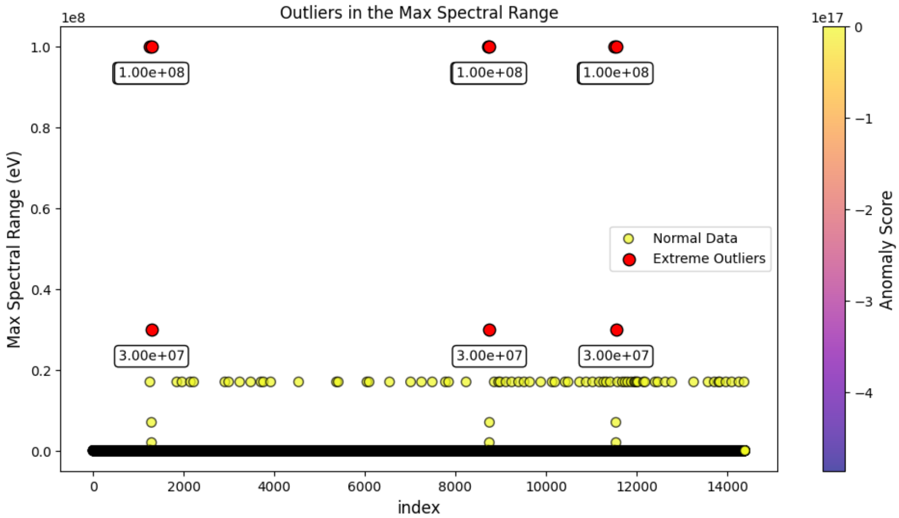

Figure 6: Local Outlier Factor Implemented with the Data

Using python's scikit-learn package (Pedregosa et al., 2011), the local outlier factor was applied to the max spectral range values. The number of local neighboring data points was set to 20 to provide a good balance between sensitivity and not including non-extreme points. The scatter plot shown above displays the maximum energy data and highlights red outlier points that are shown over the rest of the dataset. The local outlier score gradient ranged from a lower score at the yellow points to a higher score at the red points. Each of the higher energy solar flares are shown to be high above the large density of normal values. This separation indicates that these anomaly points are no normal cases, they are physically distinct events. The local outlier factor was set to return a score for each data point. This means that there is a cutoff of a score below -1.5 to mark the extreme outliers. As shown in the figure above the local outlier factor scores clearly show the separation of some high energy data points from the dense floor of normal values. These extreme outliers correspond to the solar flare energies of 30 million electron volts and 100 million electron volts respectively. Due to the fact that satellites only have a tolerance of about 8000 electron volts, it can be concluded that that majority of the satellites and GPS systems in the area of the solar flare were deeply affected (Khodairy et al., 2020)(Shigeoka et al., 2004). To validate the assumption that was mentioned earlier, the timestamps of the extreme outliers were compared to other timestamps from other papers. According to Wu, C.-C. (2016), during the March 2015 solar storm, specifically on the 17th and 18th, there was a widespread communication blackout and GPS faults (Wu, 2016). These faults were directly related to the solar storm occurring during those days(Wu, 2016). The outliers in the data perfectly align with the timing of the satellite and GPS errors which confirms the reality of what the statistics say. Therefore, it is correct to correlate the outliers detected by the local outlier factor with the real world events.

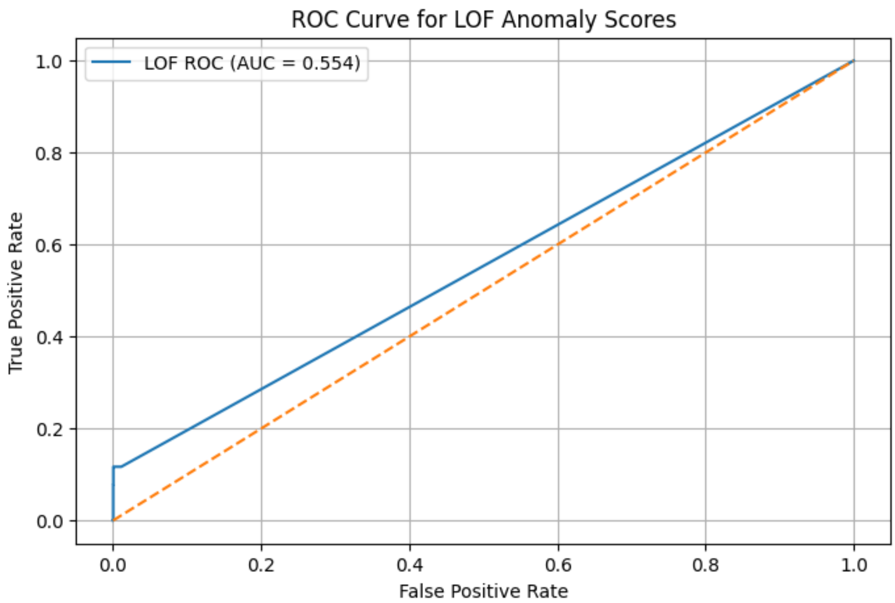

Figure 7: ROC Visualization

To determine the local outlier factors performance, the ROC test was done using binary labels (1 or 0) where the solar flare events above 10 million electron volts were configured to be anomalies. The machine learning model achieved a score of 0.554. This means that it was slightly above guessing randomly. Even though the score is not ideal, it is better than guessing randomly. With these results combined with the results from previous papers (Wu, 2016), the local outlier factor's algorithm is still considered somewhat correct even with a non ideal ROC score.

4 Conclusion

The overall goal of this study was to analyze the highest energy solar events using the local outlier factor, the unsupervised machine learning model. The focus was on identifying the outliers in the maximum spectral range from the extremely intense March 17th to 18th, 2015 solar storm. This study was aimed to answer the question of how high intensity solar flares affected satellite and GPS systems. This research question has provided technological insight into their vulnerability to high energy solar storms. The findings display a clear fact that a small, yet critical amount of high intensity solar flares emitted energies that are far beyond the critical tolerances of satellite and GPS systems (at around 8000 electron volts) (Khodairy et al., 2020). Data analysis has revealed that while there were many normal solar energy values, some values reached from 30 million to 100 million electron volts. These values were shown to be extreme using the local outlier factor. These values aligned with the values with previous reports such as Wu, C.-C. (2016) (Wu, 2016). Although the ROC score was only 0.554, the outliers it identified still corresponded with significant impact towards the satellite and GPS systems. This research contributes to the forever increasing knowledge of space weather analysis and machine learning as it demonstrates that large scale anomaly detection can be done using models such as the local outlier factor. This is due to the data's skewed points and the local outlier factor's ability to determine them as extreme points. However, this study also has limitations. The first limitation would be that the dataset is missing which solar flares could have caused significant damage. The second limitation is that only the local outlier factor was used. Other machine learning models such as isolation forest were tested but yielded results that would be considered mediocre. The last limitation would be the type of information that was contained inside the data. The data could have used more variables such as flux or magnetic field values. Future research will expand on this work that was done by exploring more advanced machine learning models. Getting more research on this problem can also help the infrastructure of satellite and GPS systems as they are crucial systems for humanity. Specifically with respect to the aviation industry, military, and emergency services. Finally, this study also shows that machine learning can play a crucial role in protecting these systems and improving humanity.

References

Gopalswamy, N. (2010). Solar storms and their influence on space weather.

Khabarova, O. (2024). Recent advances in solar wind studies. *Frontiers in Astronomy and Space Sciences*.

Khodairy, S., Sharaf, M., Awad, M., Abdel Hamed, R., and Hussein, M. (2020). Impact of solar activity on low earth orbiting satellites.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., and Duchesnay, E. (2011). Scikit-learn: Machine learning in python. Retrieved from `https://scikit-learn.org/stable/about.html#citing-scikit-learn`.

Ray, S. et al. (2016). Impact of march 2015 solar storm on satellite operations. *Space Weather Journal*.

Shigeoka, N. et al. (2004). Radiation effects on electronics in space. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 37:2303.

Virtual Solar Observatory (2025). Solar data search interface. `https://sdac.virtualsolar.org/`.

Wu, C.-C. (2016). Interplanetary shocks and their effects on earth's magnetosphere. *Earth, Planets and Space*, 68(1).