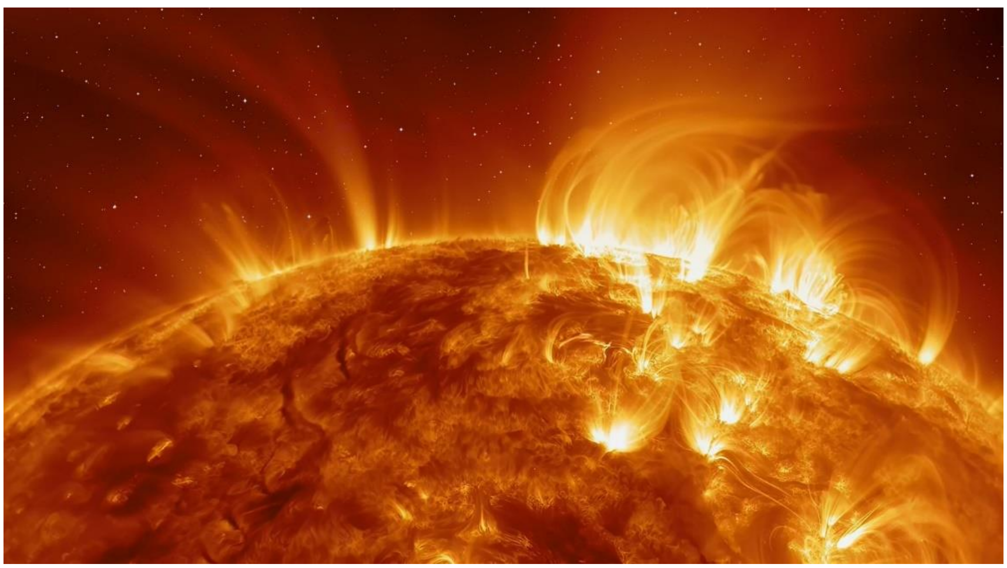
Raunak Sarbajna and Christoph F. Eick

COSC 3337 *"Data Science I"* Fall 2023
Group Project (group size 4 students)

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



Last Updated: October 20, 2023, 11a (Version 8')
½-3/4 page status report is due: Friday, October 27, 2023
All other group project deliverables are due: Friday, November 10, 2023
Responsible TA: Raunak Sabajna

Learning Objectives:
1. Summarizing complex spatio-temporal data
2. Density Estimation
3. Hotspot Discovery
4. Change Analysis

The goal of the project is to design and implement a system called *Helios*, which is capable of summary generation, mapping, hotspot discovery, and change analysis of high-intensity solar flares events.

A Solar flare is the rapid release of a large amount of energy stored in the solar atmosphere. During a flare, gas is heated to 10 to 20 million degrees Kelvin (K) and radiates soft X rays and longer-wavelength emission. Unable to penetrate the Earth's atmosphere, the X rays can only be detected from space. Instruments on Skylab, SMM, the Japanese/US Yohkoh mission and other spacecraft have recorded many flares in X rays over the last twenty years or so. Ground-based observatories have recorded the visible and radio outputs. These data form the basis of research which tries to deepen our understanding why solar flares occur and the characteristics of their occurrence.

DATA DESCRIPTION

Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI, originally High Energy Solar Spectroscopic Imager or HESSI) is a NASA solar flare observatory. It is the sixth mission in the Small Explorer program, selected in October 1997 and launched on 5 February 2002. Its primary mission is to explore the physics of particle acceleration and energy release in solar flares.

Researchers believe that much of the energy released during a flare is used to accelerate, to very high energies, electrons (emitting primarily X-rays) and protons and other ions (emitting primarily gamma rays). The new approach of the HESSI mission is to combine, for the first time, high-resolution imaging in hard X-rays and gamma rays with high-resolution spectroscopy, so that a detailed energy spectrum can be obtained at each point of the image.

ATTRIBUTES

1. *Flare* - An ID number, ymmddnn, e.g., 2042101 is the first flare found for 21-Apr-2002. These numbers are not time ordered.
2. *Date* - The date when the flare occurred.
3. *Start* - Flare start time
4. *Peak* - Flare peak time (peak counts/second in energy range 6-12 keV, averaged over active collimators, including background.)
5. *End* - Flare end time
6. *Dur[s]* - Duration of flare in seconds
7. *Peak[c/s]* - Peak count rate in corrected counts, peak counts/second
8. *Total Counts* - Total of counts in corrected counts, (counts in energy range 6-12 keV integrated over duration of flare summed over all subcollimators, including background.)
9. *Energy [keV]* - The highest energy band in which the flare was observed. Electron Kev (kilo electron volt) ['6-12', '12-25', '25-50', '50-100', '100-300', '300-800', '800-7000', '7000-20000'].
10. *X pos [asec]* - Flare position in arcsec from sun center [-1007, 1005]
11. *Y pos [asec]* - Flare position in arcsec from sun center [-998, 1012]
12. *Radial [asec]* - Radial distance in arcsec from sun center

Example data:

flare	**start.date**	**start.time**	**peak**	**end**	**duration.s**
2021213	12-02-2002	21:29:56	21:33:38	21:41:48	712
2021228	12-02-2002	21:44:08	21:45:06	21:48:56	288
2021332	13-02-2002	00:53:24	00:54:54	00:57:00	216

peak.c/s	**total.counts**	**energy.kev**	**x.pos.asec**	**y.pos.asec**	**radial**
136	167304	12-25	592	-358	692
7	9504	06-12	604	-341	694
15	11448	06-12	310	375	487

DATA USAGE

The total data will be subdivided into 2 sets, each of which contain the locations, highest energy levels and peak counts of events for a time interval of 2 years; one set will contain the data from 2004 to 2005 and the other from 2015 to 2016 and your system will use a sliding window approach to create input data for the Helios system. The sets have been chosen to be in a similar period of the 11-year solar cycle. The data has been filtered to remove the 3-6 KeV band, and radial outliers.

The coordinate system for the dataset is in arc-seconds. You can plot them without any transformation, with the knowledge that (0,0) is the center of the Sun. The diameter of the sun is 0.5 degrees which is 30 arcminutes or 1800 arcseconds. X-axis is along the horizontal axis and the Y-axis is along the vertical.

The energy band tells you which band in which the peak value fell within that flare. So if the value is 12-25, then it means that the peak solar flare emission was in the 12 KeV to 25 KeV range.

<u>Task 1: Solar Flare Intensity Estimation</u>

Use Set 1 (2004 – 2005) for this task.

We subdivide Set 1 into smaller batches, assuming a batch size of 4 months and 2 months overlap between consecutive batches; that is, there will be batches for months 1+2+3+4, months 3+4+5+6, …, months 21+22+23+24 (11 batches).

Develop two methods of flare intensity estimation; intensity estimation techniques measure the flare intensity in a location ((X,Y)) based on a set of flare events. Method 1 measures the intensity based on the **total.counts** attribute. Method 2 measures the intensity based on the **duration.s** and **energy.kev** attributes.

Subtasks of Task1:

1. Develop Method 1
2. Develop Method 2
3. Create intensity maps for months 1+2+3+4 using Method 1 and Method 2
4. Create intensity maps for months 21+22+23+24 using Method 1 and Method 2
5. Compare the 4 maps you generated; look at spatial variation and total intensity.

Remember that **energy.kev** is provided as bands, and there are 8 bands total. Creative visualization techniques will be granted more points.

Task 2: Hotspot Discovery and Analysis

Use Set 1 (2004-2005) for this task.

Develop hotspot discovery techniques for the intensity maps you generated in Task1. We assume a hotspot is a contiguous polygon[1] in a 2D X-Y space for which the event intensity of points inside the polygon is above a user-defined intensity threshold. Your system should create two kinds of hotspots:

a. Small, very hot spots whose density is above a "high" intensity threshold d1
b. Large, more regional hotspots whose intensity if a above a "medium high" intensity threshold d2; d1>d2.

Finding proper density thresholds d1 and d2 to create those two kinds of hotspots is a problem you need to solve in this project. However, in the Task2 experiments we restrict our attention to intensity analysis based on Method 1.

Subtasks of Task2:

a. Design and implement a hotspot discovery algorithm
b. Determine intensity threshold d1 and d2 for Method1, based on your results for Task1
c. Create a method to visualize hotspots
d. Create a time series of 11 hotspots for the 11 batches using intensity threshold d1
e. Create a time series of 11 hotspots for the 11 batches using intensity threshold d2
f. Summarize and interpret the results you obtained in step d
g. Summarize and interpret the results you obtained in step e
h. Very briefly also compare the results you obtained in steps d and e

Task 3: Change Analysis for Solar Flares

Here, your goal is to compare the solar flare data from Set 1 (2004 to 2005) with those from Set 2 (2015 to 2016), and to summarize the major differences between the two datasets. Compare the two kinds of intensities you analyzed before and also analyze spatial variation. Moreover, compute basic statistics (e.g. various counts and averages) for Set 1 and Set 2 and compare those.

STATUS REPORT

You will need to submit a ½ to ¾ page single-spaced status report by the end of day if Friday, October 27; the status report counts 7% toward the group project grade.

ADDITIONAL INFORMATION

The required data can be found in the *Files* tab of the *Datasets and Code* channel in the COSC-3337 Teams page. The two datasets are: Set 1: Solar_flare_RHESSI_2004_05.zip, and Set 2: Solar_flare_RHESSI_2015_16.zip.

---

[1] However, you might create hotspots of "simpler" shapes, instead of polygons; e.g. rectangular hotspots or hotspots which are contiguous regions of grid cells.

You must present your findings in a report of 8 ~ 10 pages, excluding references and appendices. Any animations created for the report must be attached alongside (preferably in .mp4 or .gif formats). The report must also contain clearly labelled imagery of the intensity, hotspots and the change-over-time.

In your report, you must clearly state which members contributed to the project, and what were their contributions. Each group must submit only one report. Zip up your report/code/animations. Name your zip as <TEAM_NAME_Group_Project_3337.zip>. Your report should be submitted in PDF format and formatted using the ACM camera-ready templates available at http://www.acm.org/publications/proceedings-template. You may use MS Word/Google Docs/LaTeX/etc. for your report. Additionally, you must also submit your entire, *working* code as a zipped folder

.