Stealth Coronal Mass Ejection Detection

Extended Abstract

Shawn Polson
University of Colorado, Boulder
USA
shawn.polson@colorado.edu

Tyler Albee
University of Colorado, Boulder
USA
tyler.albee@colorado.edu

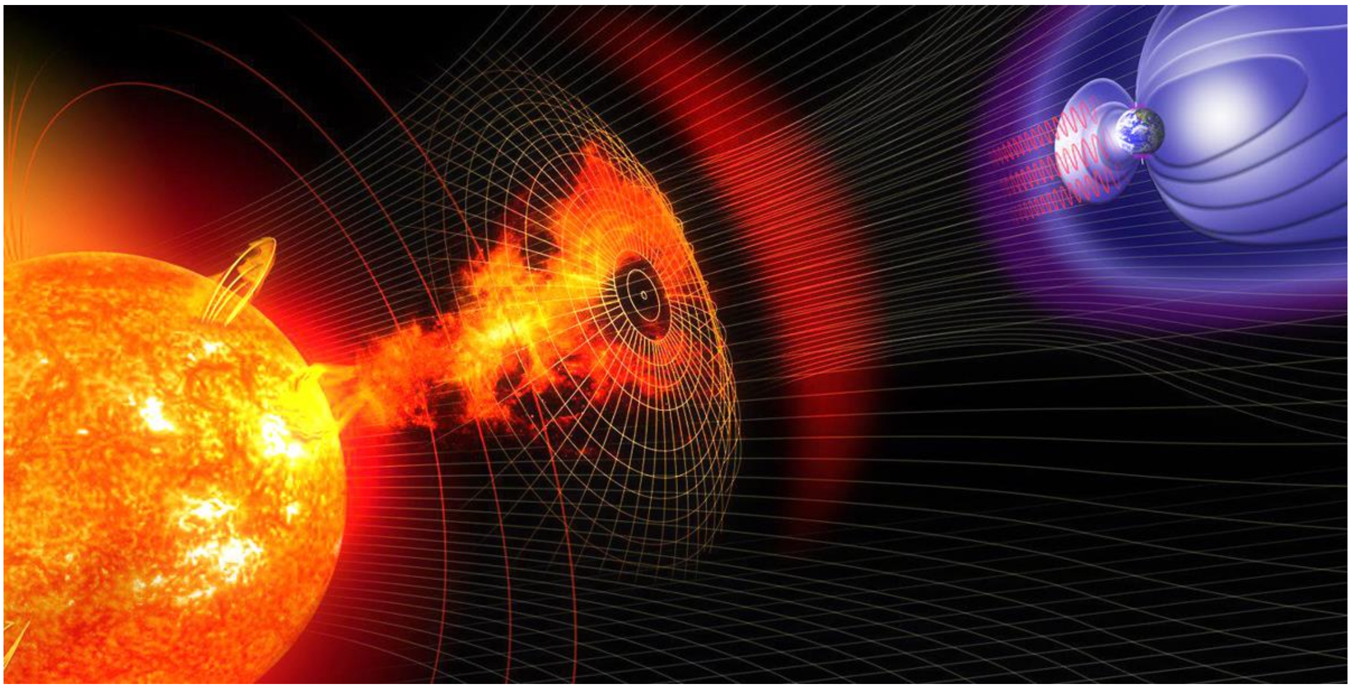

Figure 1: An artistic representation of a coronal mass ejection

ABSTRACT

This project intends to help with the detection and characterization of stealth coronal mass ejections (CMEs) by classifying coronal dimming events in the Solar Dynamics Observatory Extreme Ultraviolet Variability Experiment (SDO EVE) telemetry. Stealth CMEs are solar events not preceded by obvious solar activity wherein portions of the Sun's corona erupt into space. These are distinct from classic CMEs mainly in their lack of obvious preceding solar activity. SDO EVE is an instrument in space equipped with spectrographs and other measuring devices, and is maintained by the Laboratory for Atmospheric and Space Physics (LASP) at the University of Colorado Boulder [1].

KEYWORDS

Stealth, CME, Coronal Mass Ejection, detection, coronal dimming, LASP, SDO EVE, spectrograph, data mining

ACM Reference format:

S. Polson, T. Albee. 2018. Stealth Coronal Mass Ejection Detection. In *CSCI-4502, Boulder, Colorado USA, March 2018 (CSCI-4502)*, 3 pages https://github.com/tyleralbee/Stealth-CME-Detection.

1 PROBLEM STATEMENT/MOTIVATION

Stealth CMEs are a hot topic in space science; the mechanisms behind these mysterious solar eruptions were just uncovered last year (2017) [2]. This is largely why LASP is actively involved in stealth CME research. We became involved with this work by reaching out to scientists at LASP to see what areas of research our data mining knowledge could potentially assist. That led us to our subject matter experts (SMEs): James Mason and Don Woodraska. More information about them can be found in the "Acknowledgements" section.

In short, this project aims to create the first ever catalog of stealth CMEs by cataloging coronal dimming events—that are unrelated to classic CMEs—in the SDO EVE data. Coronal dimming events are the mark of CMEs because CMEs leave behind temporary voids when they depart from the solar corona. Information about a CME, such as its mass and velocity, can be learned by analyzing these dimming events. As such, this project also aims to characterize the dimming events in terms of slope and depth. This would extend prior work done in this field. In particular, this project will expand upon the work done by James Mason in his PhD dissertation [1], where he pioneered methods of characterizing classic CMEs by observing coronal dimming events.

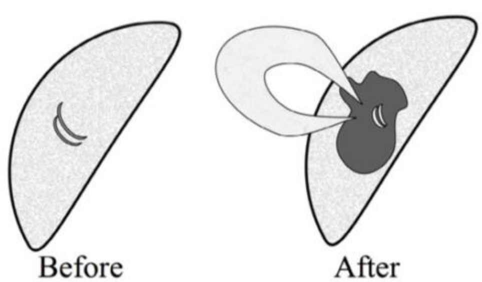

Figure 2: An illustration of a void left in the Sun's corona after a solar eruptive event

The SOHO project of the European Space Agency (ESA) and NASA maintains a catalog of CMEs at the CDAW Data Center [3], and within the scientific community, it is the go-to catalog for solar eruptive events. It currently holds information about every recorded CME from January 1996 to the present day, but it notably does not label CMEs at "stealth" or not. This project aims to label the stealth CMEs in this catalog. That would allow space scientists to finally be able to study the correct solar events to learn more about stealth CMEs. It is noteworthy that stealth CMEs differ from classic CMEs in their lack of any obvious preceding solar activity and in their lower magnitudes. Because of this, novel detection methods will have to be invented during this project. It will also be trickier to detect stealth CMEs because solar flare events (or similarly known solar activity) cannot be used to "trigger" searches for them. We are essentially trying to find a smaller needle in a larger haystack.

2 LITERATURE SURVEY

As previously mentioned, stealth CMEs are a relatively new area of research, and this project is the first of its kind. However, it will expand upon work done by James Mason in his PhD dissertation entitled "Solar Eruptive Events: Coronal Dimming and a New CubeSat Mission." In his dissertation, Mason pioneered methods of characterizing classic CMEs by observing coronal dimming events. This project will mine for data in essentially the same dataset used by Mason (SDO EVE telemetry), so his methods can serve as the lead inspiration for our methods that will ultimately detect stealth CMEs.

3 PROPOSED WORK

This project has three main goals: to detect coronal dimming events in SDO EVE emission line time series data without the use of solar flare event times to trigger the search, to characterize the detected dimming events in terms of slope and depth, and to label stealth CME events in the SOHO catalog as "stealth." To accomplish this, the authors will take two main approaches. The first approach will be to perform a sliding correlation coefficient analysis using the signature of a CME as the basis for what to look for. By this we mean that we have cleaned and analyzed the data from the time window of one noteworthy CME from a case study in James Mason's dissertation, and using a sliding time interval, we will search for intervals in our data set whose data sufficiently match our CME signature, according to calculated correlation coefficients. The second approach will be to generally look for suspicious slope drops in light curves from the SDO EVE data set, but that approach may require statistical analysis that is beyond us—we have yet to try it. We will work with our SMEs to develop and improve our detection methods.

Data cleaning will involve assigning missing values to null values (the current proposed method is to assign the value "-1" to all null values), removing unneeded attributes from our data set, and also smoothing the light curves in our data set to eliminate noise from the EVE detector. Another consideration is the fact that a power anomaly in the *MEGS-A* spectrograph (described in the "Data Set" section) rendered the instrument unable to measure wavelengths from 6nm to 33nm beyond May 26, 2014. Aside from that, no other cleaning is expected to be needed, as the SDO EVE data has already normalized the measured solar irradiance values to what they would be at a distance of 1 AU.

No data integration is planned for our data set because the SDO EVE data has all needed solar measurements. However, we will compare our findings against two data sets to evaluate them: The SOHO catalog (to identify whether time ranges we find do indeed contain CMEs) and a data set from a Geostationary Operational Environmental Satellite (GOES) instrument [1]. That data set will be used to distinguish stealth CMEs from classic ones because it records solar flare activity; no such activity around a CME means it was a stealth CME.

4 DATA SET

The data set for this project comes from SDO EVE telemetry. SDO EVE stands for "Solar Dynamics Observatory Extreme ultraviolet Variability Experiment," and it is an instrument that has been in orbit in space since 2010 [1]. Before describing the data, it is pertinent to better describe what the instrument is.

Onboard the EVE instrument are two spectrographs, known as *MEGS-A* and *MEGS-B*. Together, they measure the irradiance of the Sun's spectrum from 6nm to 106nm. *MEGS-A* failed due to a power anomaly on May 26, 2014, so after that date, only wavelengths from 33nm to 106nm are measured. Two forms of data are regularly produced by this instrument: "lines" data and "spectrum" data. The "lines" data represents the solar irradiance measurements in terms of irradiance (in W/m^2) at selected emissions lines (elements such as Fe XVIII or He II). The "spectrum" data represents solar irradiance across the whole spectrum of wavelengths. Measurements are taken with a ten second cadence for three hours of every day. Before the *MEGS-A* power anomaly, measurements were taken every ten seconds all hours of the day. The instrument is still in operation and has been in operation since February 11, 2010.

The data is provided and maintained by LASP, which is where Shawn Polson is employed.

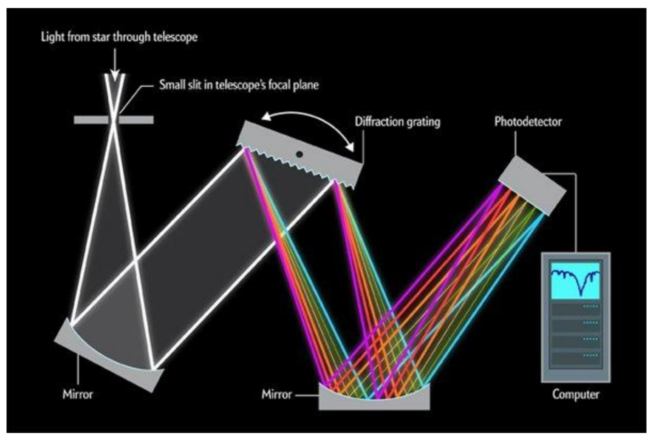

Figure 3: A rough depiction of how a spectrograph splits light into its component wavelengths

5 EVALUATION METHODS

5.1 Metrics

- **Accuracy** (**Accurate Detection Rate**) – The ADR of our detection of a stealth CME event. What percentage of the time does the software trigger a false positive? a false negative? In order to achieve the highest possible accuracy, the events will be checked against SOHO catalog of CMEs. This will be gauged largely with *recall* (TP/(TP+FN).

- **Applicability** – The team's success depends on the applicability of the gained knowledge. Finding and cataloging stealth CMEs from past data would be beneficial to the scientific community, and this project will partially be evaluated in terms of how helpful its results are. The applicability of the project can be quantified into a number on a scale of 1-10 (for example).
- **SME Approval Rating** – The project is under supervision by two subject matter experts, Don Woodraska and James Mason. The feedback given to us by Don and James on our work can be quantified into a number on a scale of 1-10 (for example).

5.2 Existing Solutions

James Mason has devised an existing solution for the detection and categorization of classic (non-stealth) CMEs using Python. These CMEs detected by Mason's code have been cross checked with pictures to prove that the dimming events in the spectrograph data are indeed coronal mass ejections. Recall that a *stealth* CME has no obvious preceding solar activity in the corona, but still results in a CME-like dip in the spectrograph data. The logic in Mason's software to detect CMEs in the data can thus be used as a starting point for the logic of detecting stealth CMEs.

6 TOOLS

- **Python** – Python will be the language used to compile data into graphs and perform analysis. The version used throughout the project will be Python 3.6.
- **IDL** – IDL will be the last resort in cleaning, preprocessing, and analysis. The FITS files and the saveset given to us are native to IDL. If we have problems with Python, we will use IDL.
- **LaTiS** – LaTiS is a data modeling framework developed and maintained at LASP (partially by Shawn Polson). It employs a "functional data model" which captures the functional relationships between attributes, and will serve as an alternative to Python or IDL. LaTiS will also serve as a more powerful file conversion tool; when working with FITS files, it is often advantageous to convert them to CSV files.
- **Pandas** – The pandas library will be used for the creation and manipulation of dataframes within Python.
- **Matplotlib** – The generation of graphs and charts for interpretation will be done using the matplotlib library in Python.
- **Numpy** – numpy will be used for the high-level manipulation of numbers within the dataframe
- **savReaderWriter** – savReaderWriter is used to take the saveset originally created in IDL and format it to be read into a Pandas dataframe in Python.
- **Jupyter** – Jupyter Notebook will allow for code, writing, and pictures to be cleanly interwoven. The technical presentation of findings will be most efficiently shown in Jupyter Notebook.

7 MILESTONES

7.1 Shawn - Programming Milestones

0. Optimize LaTiS to be accepting of FITS format files
1. Get data imported and readable in LaTiS
2. Clean data of NULL values in LaTiS
3. Experiment with the usefulness of LaTiS for data analysis
4. Compile the data into another saveset for future work
5. Use LaTiS to quickly and efficiently convert between FITS and CSV files
6. Clean saveset and determine characteristics of our sample CME using James Mason's Python routines
7. Work with Tyler to perform the correlation coefficient analysis as described above

8. Work with Tyler to evaluate findings against SOHO catalog and GOES data as described above.

7.2 Tyler - Programming Milestones

0. Take the IDL saveset into Python and read it
1. Experiment with the plotting and visualization of the data in Python
2. Become acquainted with James' functions and methodology in Python
3. Make changes to James' functions and methodology in Python based upon theory developed during this project
4. Comment changes to James' functions and methodology in Python
5. Finalize newly-written Python routines
6. Work with Shawn to perform the correlation coefficient analysis as described above
7. Work with Shawn to evaluate findings against SOHO catalog and GOES data as described above
8. Tweak correlation coefficient methodology based on findings
9. Get extensive peer review on new code
10. Present new code/results to James Mason and Don Woodraska for professional evaluation

7.3 Milestones Completed

We were able to compile the raw SDO EVE data into a usable saveset. This involved converting it into a usable format and using pandas in Python to remove all attributes from the data set except for the time attribute and the measured solar irradiance values. Then, using pandas in conjunction with a light curve-fitting routine written by James Mason (which itself uses a chi-squared test to do the fitting), we were able to preprocess our compiled saveset; i.e., remove instrument noise from it. An example light curve fitting is shown in Figure 4.

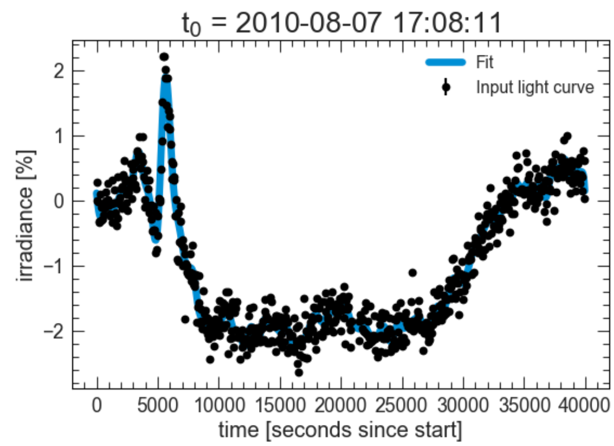

Figure 4: An example output plot of irradiance values that have been fitted to a light curve. Note the initial spike and the proceeding dimming which is characteristic of a CME.

We also successfully got the "signature" of a CME, which will be used to perform our correlation coefficient analysis. This involved sampling our data set into only the samples within the time range of our CME, converting the raw irradiance values from each of the 39 light curves for each emission line into percent differences from our start time, fitting each of those light curves to reduce noise, and using three Python routines written by Mason to get the dimming slope, depth, and duration of our CME. These results are detailed in the "Results" section.

7.4 Milestones Todo

We now have the signature of a CME and data to work with. What remains to be done, then, is to find as many stealth CMEs as possible in our data. As described above, we will first attempt this by performing a sliding correlation coefficient analysis. Using a sliding time interval, we will search for intervals in our data set whose data sufficiently match the data of our CME signature—according to calculated correlation coefficients. Once we have those identified time intervals, we will be holding a set of *potential* CME events. Our next task will then be to check those time ranges against the SOHO catalog to confirm whether they are in fact CMEs, and to check them against the GOES data to see if solar flare activity was recorded around each time interval.

If an identified time interval has a matching CME in the SOHO catalog and if it is not associated with GOES flare activity, we can be sure it is a stealth CME. If we are lucky, our first pass will turn up *some* stealth CMEs.

Regardless of the results of our first pass, we will assess the accuracy of our correlation coefficient analysis (using *recall* among other methods), and then we will work with our SMEs to refine our methodology. We hope to eventually end up with software that can identify every stealth CME in our data set, but that is a lofty goal and perhaps not feasible. Finally, we will characterize whatever stealth CMEs we do find in terms of their dimming depth, duration, and slope.

8 RESULTS SO FAR

Our biggest achievement to date was using Python to preprocess our data and to generate the signature of a CME. In doing this, we got ourselves over the learning curve of this project (e.g., hurdles related to our scientific understanding), and our methods will assist us in finding and characterizing stealth CMEs. We have software that can clean our data and can characterize CMEs in terms of their dimming depth, duration, and slope; this is pictured in Figures 5, 6, and 7, respectively. Figure 8 shows all of these characteristics combined into one plot. We will eventually characterize all discovered stealth CMEs according to these metrics.

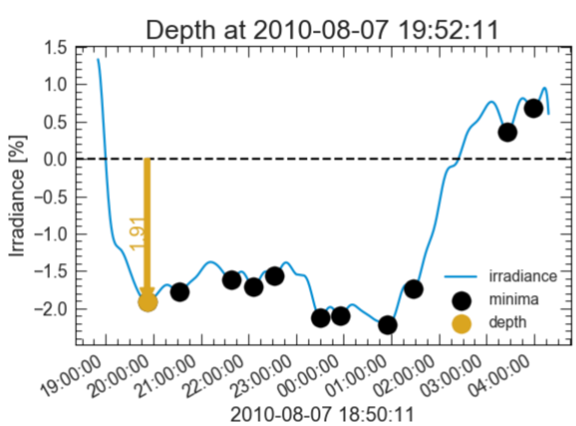

Figure 5: A plot showing the detected dimming depth of a light curve during a CME

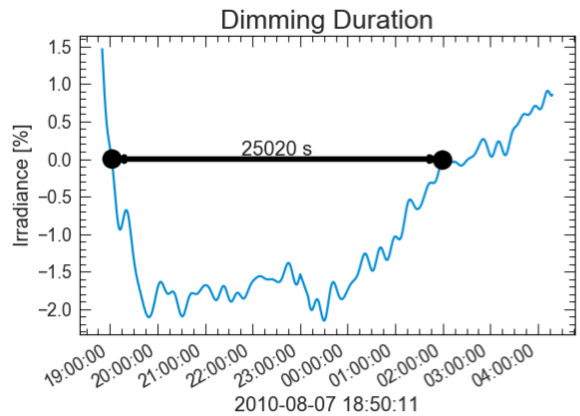

Figure 6: A plot showing the detected dimming duration of a light curve during a CME

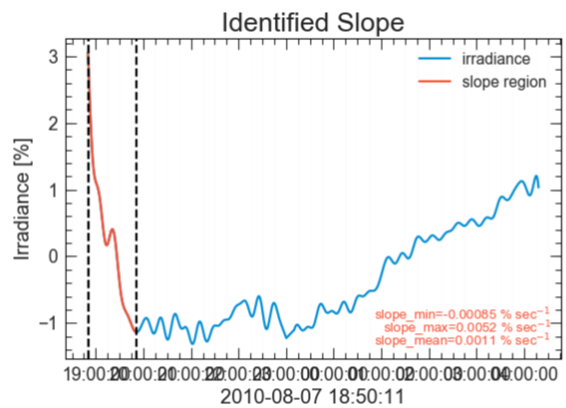

Figure 7: A plot showing the detected dimming slope of a light curve during a CME

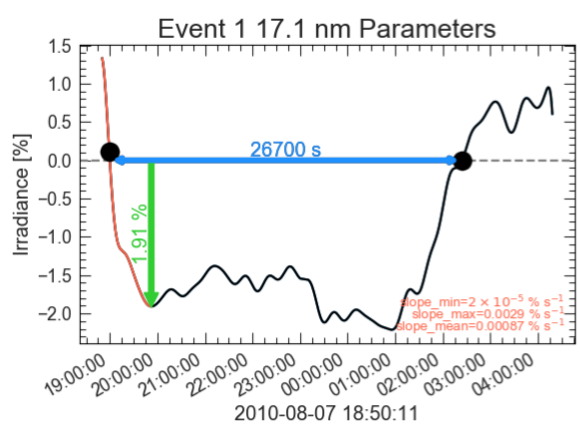

Figure 8: A summary plot showing all detected characteristics of a light curve during a CME: dimming depth, duration, and slope

ACKNOWLEDGMENTS

First on the list to thank is James Mason, who wrote his dissertation on CMEs and coronal dimming. James Mason is currently a post-doc at NASA's Goddard Space Flight Center. He provided the knowledge and inspiration for this project. Don Woodraska, another SME who works at LASP, deserves a massive thank you for the coordination and the knowledge he has provided us. Work done in LaTiS was made possible by the LASP Web Team, with special consideration to Doug Lindholm, the man who envisioned and created LaTiS back in the 1990s. We are honored to be working among these extremely talented individuals.

REFERENCES

[1] Woods, Tom et al. "Instrument." Laboratory for Atmospheric and Space Physics, lasp.colorado.edu/home/eve/science/instrument/.

[2] Grossman, David. "Scientists Figure Out the Sun's Mysterious 'Stealth' Explosions." Popular Mechanics, Popular Mechanics, 10 May 2017, www.popularmechanics.com/space/solar-system/news/a26462/sun-stealth-cme/.

[3] Author, Electronic. "CME List." *NASA*, NASA, 25 May 2014, cdaw.gsfc.nasa.gov/CME_list/UNIVERSAL/