Stealth Coronal Mass Ejection Detection

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

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

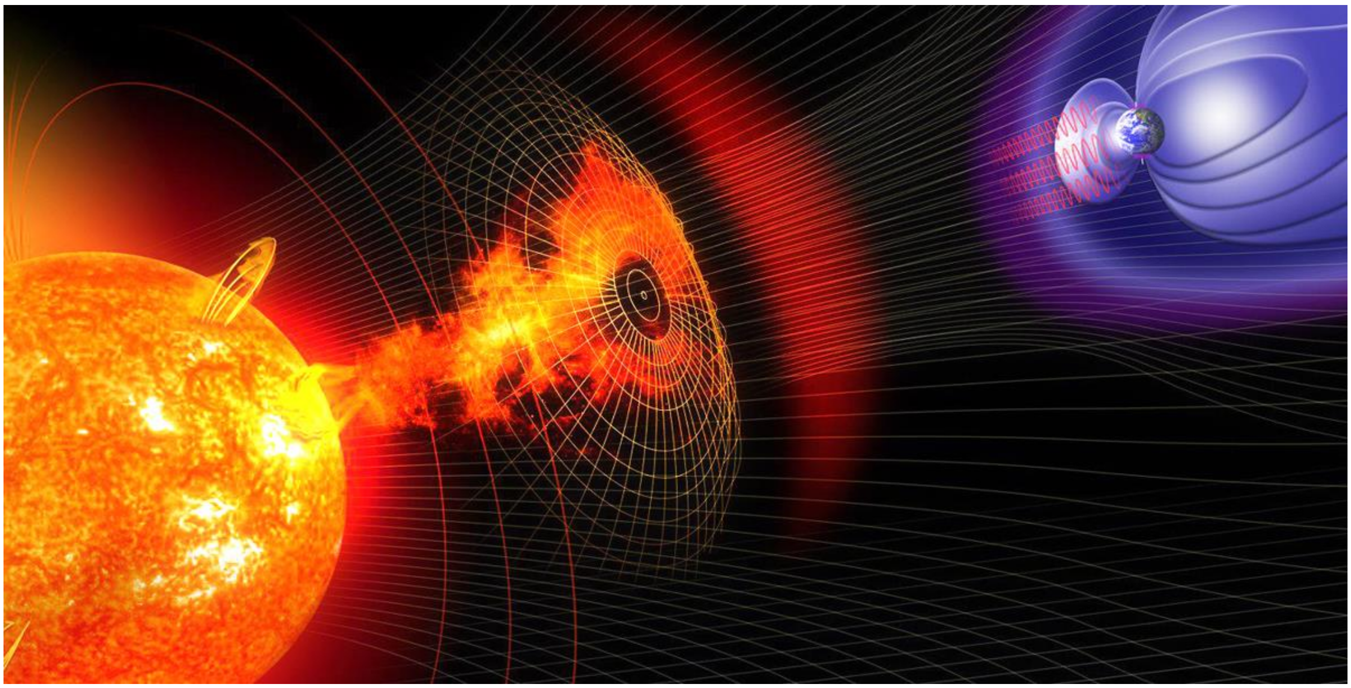

Figure 1: An artistic representation of a coronal mass ejection

ABSTRACT

This project aimed to help with the detection and characterization of stealth coronal mass ejections (CMEs) by characterizing 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]. While we did not successfully characterize stealth CMEs in this project, we have detected and labeled potentially many of them—further study is needed—and we have written a collection of Python routines that would indeed characterize them, if given enough time to run. Our research suggests that around 32.1% of our sun's CMEs are stealth CMEs.

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 INTRODUCTION

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 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 aimed 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. We were seeking to answer questions like "What proportion of our sun's CMEs are stealth CMEs?" "How can stealth CMEs be characterized by their dimming profiles?" and "What is the risk of a stealth CME damaging Earth's technology?" That last question is of particular importance because CMEs are events that are in fact capable of causing great calamity and destruction on Earth if charged enough and pointed in our direction. And stealth CMEs (while typically smaller in magnitude) give no warnings like their classic counterparts.

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 aimed to characterize the dimming events in terms of their depth, duration, and slope. This extended prior work done in this field. In particular, it expanded 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.

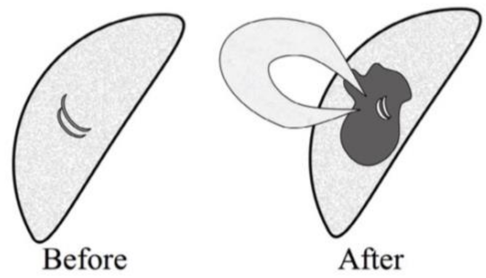

Figure 2: An illustration of a void left in the Sun's corona after a solar eruptive event

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 April 2017, but it notably does not label CMEs at "stealth" or not. This project aimed to label the stealth CMEs in that 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, a novel detection method was invented during this project. It was also trickier to detect stealth CMEs because solar flare events (or similarly known solar activity) could not be used to "trigger" searches for them. We were essentially trying to find a smaller needle in a larger haystack.

2 RELATED WORK

As previously mentioned, stealth CMEs are a relatively new area of research, and this project was the first of its kind. However, it did 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 mined for data in the same data sets used by Mason (SDO EVE telemetry and the SOHO catalog), so his methods served as the lead inspiration for ours. In writing his dissertation, Mason created a small library of software routines that perform various scientific functions, like fitting light curves to reduce noise and determining dimming characteristics of light curves such as their dimming duration, depth, and slope. This code is posted in a personal GitHub repository of his, and we were given access to it so that we could use and tweak some of his methods.

In addition to Mason's work, we worked on a problem similar to that of one other paper from a group of CME researchers: "On the Temporal Relationship between Coronal Mass Ejections and Flares" by J. Zhang et al. [4]. As part of our research, we were searching for solar flares that may had happened near CME events. As it turns out, it is exceptionally difficult to define what constitutes "near" an event. The work of J. Zhang et al. did not arrive at one well-defined answer to this question, and neither did we, so the results of our research are only as accurate as how well we defined this parameter—which left more to be desired.

3 DATA SETS

Two data sets were used in this project: a data set of SDO EVE telemetry and the SOHO catalog of CMEs. We will first explain the SDO EVE data. 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. This project used only the "lines" data. 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. Our saveset of the data contains over two million entries.

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

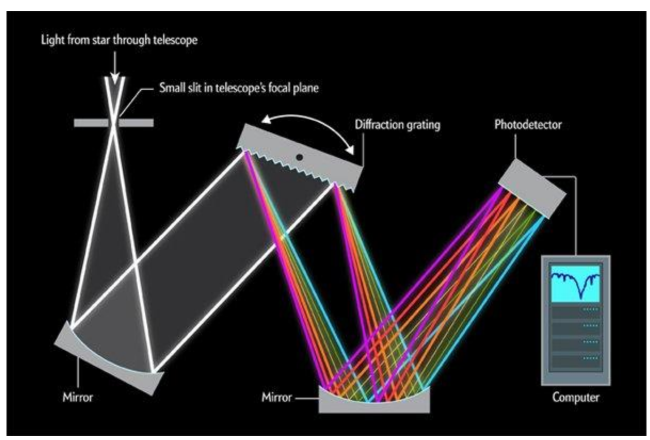

Figure 3: A rough depiction of how a spectrograph splits light into its component wavelengths

The SOHO project of the European Space Agency (ESA) and NASA maintains the SOHO catalog of CMEs at the CDAW Data Center [3]. It currently holds information about every recorded CME from January 1996 to April 2017, but it notably does not label CMEs at “stealth” or not. It does contain time labels for each event, however, and we used these known times in our research. One of our ultimate goals was to definitively label each event in this catalog as “stealth” or not. We did produce such labels, but we were too uncertain of our results to actually urge ESA and NASA to include them in the real catalog.

4 MAIN TECHNIQUES APPLIED

As stated in the introduction section, we sought to answer three main questions: “What proportion of our sun’s CMEs are stealth CMEs?” “How can stealth CMEs be characterized by their dimming profiles?” and “What is the risk of a stealth CME damaging Earth’s technology?” We developed a couple different approaches in our attempts to answer them, and we detail those approaches here.

4.1 Data Preprocessing

It was necessary to preprocess all our SDO EVE data before we could use it for any meaningful analysis. The SDO EVE data starts as raw measurements from the spacecraft itself, and because the spacecraft is in an elliptical orbit in space, the first step was to normalize all the values to what they would be at a constant distance from the sun. Luckily, because this data is provided by LASP, a world-class research organization, they were able to provide us with what they call a “level 2 revision” of the SDO EVE data, which normalizes all the raw solar irradiance values to what they would be at a distance of 1 AU. The next step was to clean the data of all null values, and we did this by assigning the global constant “-1” to each null value. This made sense because solar irradiance is always a positive value. The SDO EVE data set also contained many more attributes than we needed for our analyses, so we removed all the unneeded attributes, leaving us with a skinnier data set that was more efficient to work with—we only kept the irradiance values for each emission line and their associated uncertainties. When examining dimming events in the SDO EVE data, it was necessary to convert each irradiance value into a percentage value, because dimming is measured in terms of percent differences from a defined start time; the start times in our case were the beginnings of each CME event.

Finally, once the data had been normalized, cleared of null values, and reduced, we had to remove noise from the data, because spectrographs are noisy instruments. This cannot be done upfront, however; noise reduction must happen individually for each CME event. This means that we built noise reduction only into our software routines that characterized CMEs by their dimming profiles. For each CME event, we passed the data (as percents) into one of James Mason’s light curve fitting routines which uses a chi-squared test (χ^2) to fit the data points to a smooth curve. An example of this can be seen in the Figure 4. Fitting light curves is a calculation-heavy process, however, so removing noise ended up being the main bottleneck in our work.

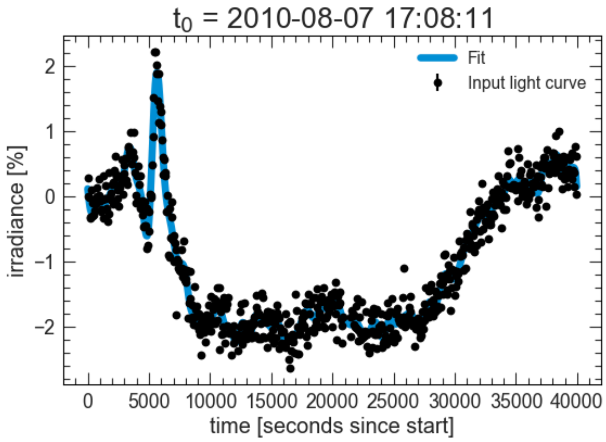

Figure 4: An example output plot of irradiance values that have been fitted to a light curve. Note the initial spike and the proceeding dimming which is characteristic of a CME.

4.2 Characterizing CME Dimming Profiles

One of the biggest hurdles we faced in this project was gaining a scientific understanding of the work that needed to be done. We picked up the work of a NASA Goddard Spaceflight Center postdoctoral researcher and essentially just ran with it, having little background knowledge on the subject, which meant we spent a lot of time learning from our SMEs. To characterize large swaths of stealth CMEs, we first had to learned how to characterize one classic one. We formalized our understanding of the material when we successfully obtained the "signature" of a CME. Specifically, we picked a CME that was the focus of a case study in Mason's dissertation, we found the data for it in the SDO EVE saveset, and we wrote a Python routine to process it. Our routine converted the irradiances to percents, enumerated all the emission lines and removed noise from each one (as described above), and then passed the preprocessed light curves into functions (written by Mason) that gave us the dimming depth, duration, and slope for that event. Example output plots from this are shown in Figures 5, 6, 7, and 8. The signature of any CME—from the viewpoint of the data—is an initial spike in irradiance, followed by a period of dimming in emission lines that are susceptible to dimming (17.1nm, 17.7nm, 18.0nm, 19.5nm, 20.2nm, and 21.1nm), followed by a return to normal irradiance values some hours after the initial dimming.

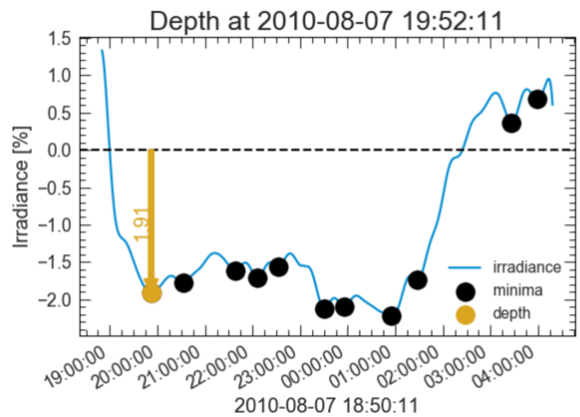

Figure 5: A plot showing the detected dimming depth of a light curve during a CME

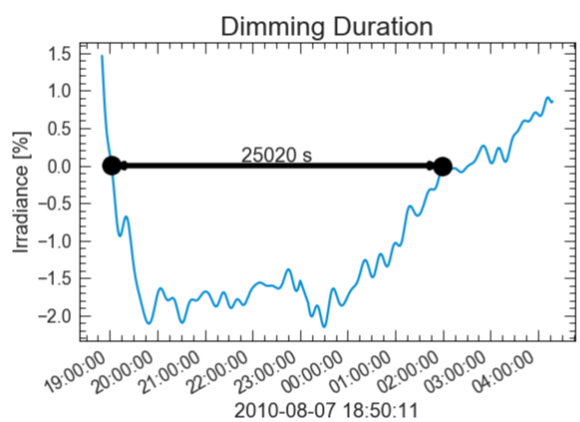

Figure 6: A plot showing the detected dimming duration of a light curve during a CME

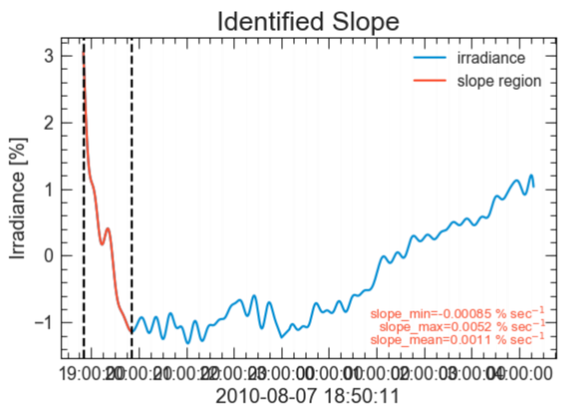

Figure 7: A plot showing the detected dimming slope of a light curve during a CME

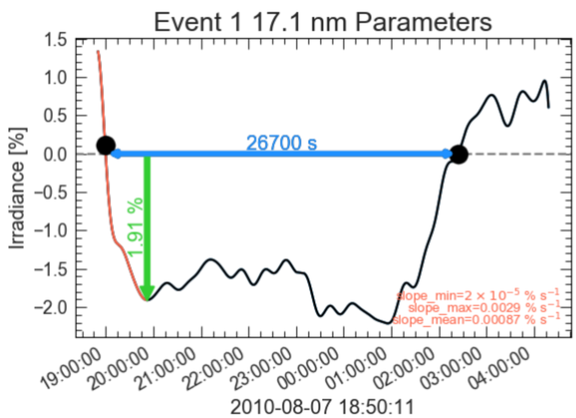

Figure 8: A summary plot showing all detected characteristics of a light curve during a CME: dimming depth, duration, and slope

4.3 Detecting Stealth CMEs (Approach #1)

Classic CMEs are intimately associated with solar flares; one can reliably use solar flare times to "trigger" searches for CMEs and that method will find most classic CMEs that occur. Stealth CMEs, on the other hand, are by definition not associated with solar flares, so flares cannot be used to trigger searches for them. Because of that fact, we had to devise a novel detection method to find the stealth CMEs we were looking for. We worked closely with Don Woodraska, our SME from LASP, and came up with the idea of combining a sliding time window with correlation coefficients to find stealth CMEs. The idea goes as follows:

The signature of a CME (as described at the end of section 4.2) is a predictable pattern in select emission lines. The signature of the CME we got from the methods detailed in section 4.2 were deemed by Woodraska to be representative of CMEs in general—that is part of the reason why Mason chose it as a case study in his dissertation—so our approach to finding CMEs was to find areas of the data which matched our CME signature. And if we could prove the match was indeed a CME and if there was no solar flare activity near the time of the event, we could be sure it was a stealth CME.

First, to find areas of the data that appeared to be CMEs, we set up a sliding time window in which to calculate correlation coefficients. The time window was as wide as our representative CME, and we slid it forward one hour after each calculation. Calculating correlation coefficients was straightforward; we were comparing samples of the same data set so the attributes (the six emission lines) already matched up, meaning that for each emission line in the time window, we fitted its light curve and calculated the correlation coefficient between it and the matching curve from our preprocessed CME, giving us six coefficients for each window, and we added them up. This means that for each time window, we would get a number between 0 and 6 telling us how well it matched the signature of our CME. We decided that a 70% match was good enough to consider it a potential stealth CME candidate, so in our Python routine, each time we find a correlation coefficient greater than or equal to 4.2, we store that time range for further analysis.

The trick, then, was to determine if each time range did in fact contain a CME and if there was solar flare activity near it. The if/then logic can be summarized as follows:

- If there is a CME in the SOHO catalog within the discovered time window
    - If there is no solar flare near the time of the event
        - Then stealth CME
    - If there is a solar flare near the time of the event
        - Then classic CME
- If there is not a CME in the SOHO catalog within the discovered time window
    - Then not a CME at all

In theory, this method will find many (if not most) stealth CMEs contained in the SDO EVE data. In practice, however, the method breaks down because (as stated in the Related Works section) it is exceptionally difficult to define what constitutes "near" an event when looking at solar flare activity. Additionally, we were unable to get around the bottleneck of doing the chi-squared tests to fit the light curves to reduce noise. Unfortunately, it takes on the order of 40 minutes to fit one set of light curves for one time window (when using our MacBook Pros), and since there were over 3,000 time windows to scan, the code we have written would take close to three months to complete. That fact did make this approach impossible to fully pursue during the project.

4.4 Detecting Stealth CMEs (Approach #2)

Upon realizing that we did not have enough time to run the correlation coefficient search, we went back to the drawing board to come up with an idea that we could execute given our time constraints. We spent an unfortunately long time developing our first approach, so our time constraints were significant at that point. After much thought, however, we had a "eureka!" moment. It was Tyler Albee who realized that, condensing the if/then logic in our previous approach, we could cut out the correlation coefficient calculating altogether and simply detect stealth CMEs by comparing the times in the SOHO catalog to solar flare events. The revised logic goes as follows:

- For each event start time in the SOHO catalog:
    - If there is no solar flare near the time of the event
        - Then stealth CME
    - If there is a solar flare near the time of the event
        - Then classic CME

Looking back, it is almost comical that neither we nor our SMEs came up with this approach sooner. It seems obvious when looking at it now.
The trick with this approach, then, was to define what constitutes "near" an event. The SOHO catalog only contains one time for each CME, namely, the start time. So, we had to decide how best to build a time range around each start time. The Python library *sunpy* has a function that, given a time range, returns a list of solar flares within it. We did not know how best to build a time range around each start time (see the end of the Related Work section), so we tried multiple different time ranges and did our best to use them to limit the uncertainties of our results.
Specifically, we ran this search with five different time range rules: three hours before and after each SOHO CME start time (six-hour windows with the start times in the middle), four hours before and after, five hours before and after, twelve hours before and after, and four hours before and two hours after (that one was the best time range from the J. Zhang et al paper). Adjusting the time windows did affect the number of CMEs that were determined to be stealth CMEs. Because of this, we do not have a definitive count of how many stealth CMEs exist

in the data, but we do have upper and lower bounds.

We did our best to use *recall* to limit the uncertainty in our results, but recall alone could not give us the confidence we needed. For clarification, *recall* is the calculation of the number of true positives (TP) over the sum of TP's and false negatives (FP): *r=(TP/(TP+FN)*. Here, a TP is a stealth CME that showed up in both searches with the narrowest time window and searches with the widest one, and a FN is a CME that showed up as "not stealth" in the wider searches but became "stealth" in narrower ones.

5 KEY RESULTS

The key result of our work is our estimation of what percentage of the Sun's CMEs are stealth CMEs. We estimate that somewhere around 32.1% of them are stealth CMEs. We were not successful in characterizing these stealth CMEs in terms of their dimming depth, duration, and slope, which was one of our main goals. But we have written a collection of Python routines that would do just that, if given enough time to run. This was an ambitious project from the start. We are not fully satisfied with our results, but our true goal was just to work with real scientists in the field, to learn their methods, and to do our best to advance their work. We can proudly say that we accomplished that goal. When we first reached out to Don Woodraska at LASP, he warned us that we were taking on an exceptionally challenging project, and he said upfront that we probably had an 80% chance of succeeding. In the end, we claim that we were 80% successful. During the course of this project, we gained an in-depth understanding of the data mining techniques used to accomplish real space science. We applied those techniques, and we came up with probably the world's first approximation of how frequently stealth CMEs occur. This was real-world experience that will stay with us forever, and we were honored to have had the opportunity to work so closely with such accomplished people.

6 APPLICATIONS

Because we were not as successful as we had hoped, the biggest application of our work will be letting scientists at LASP (and James Mason at Goddard) study the potential stealth CMEs we have labeled. Scientists have been using the SOHO catalog of CMEs since the 1990s, but the catalog has never had that "Stealth?" column. Our SMEs tell us that everyone wants that column, and our work moved that desire one step closer to reality. We did not have the time (or powerful enough machines) to characterize the potential stealth CMEs we found, but we have written Python routines that would do that if given enough time to run. It is likely that either Don Woodraska or James Mason will find a way to run our routines to get those results.

There is a saying in science that everyone stands on the shoulders of giants. Science is a collaborative and ongoing process, and it takes incremental steps to achieve anything. Future works by other researchers will improve our findings, and people will eventually have a solid understanding of our sun's stealth CME activity. We cannot speak about direct applications of our work, because there are not many. But we can speak about the benefits of understanding this solar activity in general. CMEs are a class of phenomena that, while not often thought about by the public, could potentially end modern society as we know it. If a powerful enough CME were pointed at Earth, it could fry electronics at a global scale, and the consequences of that hardly need explaining. Stealth CMEs, by nature, are not preceded by any obvious solar activity and they therefore happen out of nowhere. The risk of Earth being surprised by a nasty stealth CME is currently unknown, and since humans are fortunate enough to have a space program, we should dedicate some effort towards understanding that risk—as we did here.

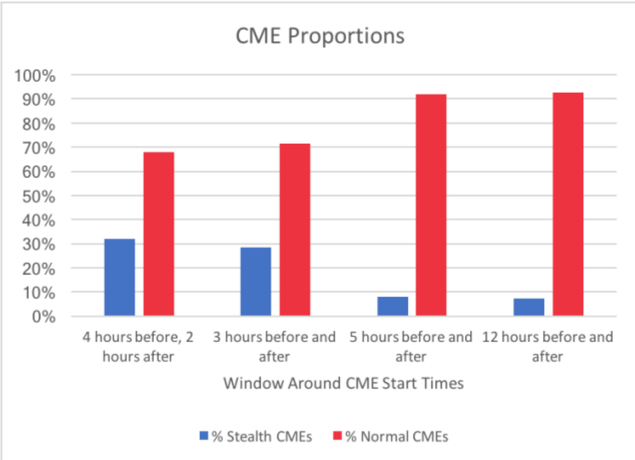

Figure 9: Proportions of stealth CMEs and classic CMEs, clustered by time windows. It is assumed that the shift in proportions for the two wider time windows are due to event overlap; the wider windows would pick up solar flares related to other CMEs.

ACKNOWLEDGMENTS

First on the list to thank is James Mason, who wrote his dissertation on CMEs and coronal dimming. James Mason is currently a postdoctoral researcher at NASA's Goddard Space Flight Center. He provided the knowledge and inspiration for this project, and worked closely with us throughout it. Don Woodraska, another SME, deserves a massive thank you for the coordination and the knowledge he provided us. He is a professional researcher at LASP, and he was our original point of contact when starting this project. He put us in touch with James Mason. 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 have worked 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/

[4] J. Zhang et al. "On the Temporal Relationship between Coronal Mass Ejections and Flares." *The Astrophysical Journal*, vol. 559, 1. 2001.