

# *Tracking CMEs using data from the Solar Stormwatch project; observing deflections and other properties*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Jones, S. R., Barnard, L. A. ORCID: <https://orcid.org/0000-0001-9876-4612>, Scott, C. J. ORCID: <https://orcid.org/0000-0001-6411-5649>, Owens, M. J. ORCID: <https://orcid.org/0000-0003-2061-2453> and Wilkinson, J. (2017) Tracking CMEs using data from the Solar Stormwatch project; observing deflections and other properties. *Space Weather*, 15 (9). pp. 1125-1140. ISSN 1542-7390 doi: 10.1002/2017sw001640 Available at <https://centaur.reading.ac.uk/71932/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1002/2017sw001640>

Publisher: American Geophysical Union

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

## RESEARCH ARTICLE

10.1002/2017SW001640

## Key Points:

- Solar Stormwatch generated a CME list tracking events through the COR1, COR2, and EUVI SECCHI imagers
- The CMEs show a systematic acceleration and growth between the COR1 and COR2 imagers
- In general, the CMEs appear to deflect toward the heliospheric current sheet

## Correspondence to:

S. R. Jones,  
shannon.jones@reading.ac.uk

## Citation:

Jones, S. R., L. A. Barnard, C. J. Scott, M. J. Owens, and J. Wilkinson (2017), Tracking CMEs using data from the Solar Stormwatch project; observing deflections and other properties, *Space Weather*, 15, 1125–1140, doi:10.1002/2017SW001640.

Received 4 APR 2017

Accepted 25 JUL 2017

Accepted article online 2 AUG 2017

Published online 5 SEP 2017

©2017. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## Tracking CMEs using data from the Solar Stormwatch project; observing deflections and other properties

Shannon R. Jones<sup>1</sup>, Luke A. Barnard<sup>1</sup>, Christopher J. Scott<sup>1</sup>, Mathew J. Owens<sup>1</sup>, and Julia Wilkinson<sup>2</sup>
<sup>1</sup>Department of Meteorology, University of Reading, Reading, UK, <sup>2</sup>Zooniverse, c/o Astrophysics Department, University of Oxford, Oxford, UK

**Abstract** With increasing technological dependence, society is becoming ever more affected by changes in the near-Earth space environment caused by space weather. The primary driver of these hazards are coronal mass ejections (CMEs). Solar Stormwatch is a citizen science project in which volunteers participated in several activities which characterized CMEs in the remote sensing images from the Sun Earth Connection Coronal and Heliospheric Investigation (SECCHI) instrument package on the twin STEREO spacecraft. Here we analyze the results of the “Track-it-back” activity, in which CMEs were tracked back through the COR1, COR2, and EUVI images. Analysis of the COR1, COR2, and EUVI data together allows CMEs to be studied consistently throughout the whole field of view spanned by these instruments (out to 15  $R_S$ ). A total of 4783 volunteers took part in this activity, creating a data set containing 23,801 estimates of CME timing, location, and size. We used these data to produce a catalogue of 41 CMEs, which is the first to consistently track CMEs through each of these instruments. We assess how the CME speeds, propagation directions, and widths vary as the CMEs propagate through the fields of view of the different imagers. In particular, we compare the observed CME deflections between the COR1 and COR2 fields of view to the separation between the CME source region and the heliospheric current sheet (HCS), demonstrating that in general, these CMEs appear to deflect toward the HCS, consistent with other modeling studies of CME propagation.

**Plain Language Summary** Coronal mass ejections are the main driver of hazardous space weather. The Solar Stormwatch citizen science project asked members of the public to find and measure coronal mass ejections (CMEs) in images of the Sun’s atmosphere taken by NASA’s twin STEREO spacecraft. In particular, participants tracked CMEs through the fields of view of three imagers that monitor different regions of the Sun’s atmosphere: an extreme ultraviolet camera, EUVI, and two Coronagraphs, COR1 and COR2. In this work we process the measurements of the citizen scientists to produce a catalogue of 41 CMEs, including details of the CME source location, size, and speed. The resulting catalogue is the first that self-consistently tracks CMEs through each of the EUVI, COR1, and COR2 imagers. We demonstrate that CMEs tend to accelerate, increase in width, and deflect toward the heliospheric current sheet as they propagate through the combined EUVI, COR1, and COR2 fields of view.

## 1. Introduction

Coronal mass ejections (CMEs) are eruptions of plasma from the Sun which propagate radially outward into the heliosphere [Gopalswamy *et al.*, 2009]. CMEs result from the release of mass and energy stored in the solar coronal magnetic field and typically take between 1 and 3 days to reach the Earth, where they can cause low-latitude auroras and geomagnetic storms [Schrijver and Siscoe, 2009]. Unfortunately, our vulnerability to these storms is increasing due to our growing reliance on technology affected by space weather [Hapgood, 2011]. Possible consequences of extreme space weather were recently discussed in a report by Cannon *et al.* [2013] and include damage and degradation of transformer cores in the electricity grid, loss of high-frequency radio communications, and anomalies in satellites from increased surface and internal charging.

Observations of the Sun are made using various different space-based and ground-based instrumentation packages. These include, but are not limited to, extreme ultraviolet imagers, coronagraphs, and heliospheric imagers, which have been widely used to study CME initiation and evolution. In 2006, NASA launched its

two identical Solar Terrestrial Relations Observatory (STEREO) spacecraft into Earth-like heliocentric orbits [Driesman *et al.*, 2007]. The STEREO “Ahead” spacecraft (STA) was launched into an orbit slightly nearer the Sun than the Earth’s orbit, and STEREO “Behind” spacecraft (STB) into an orbit slightly farther away. Consequently, STA drifted ahead of the Earth, and STB drifted behind the Earth, each by approximately  $20^\circ$  per year. The STEREO spacecraft provided the first three-dimensional view of CMEs, as throughout much of the STEREO mission phase they observed the same CMEs from different perspectives. Both STEREO spacecraft carry the Sun Earth Connection Coronal and Heliospheric Investigation (SECCHI) package, which contains two heliospheric imagers (HI1 and HI2), two coronagraphs (COR1 and COR2), and an extreme ultraviolet imager (EUVI) [Howard *et al.*, 2008]. The EUVI instrument observes the solar disk in four different wavelengths, corresponding to the emission lines of He II (30.4 nm), Fe IX (17.1 nm), Fe XII (19.5 nm), and Fe XV (28.4 nm). COR1 observes from  $1.5$  to  $4 R_\odot$ ; COR2 observes from  $2.5$  to  $15 R_\odot$ ; H1 observes from  $15$  to  $85 R_\odot$ ; and H2 observes from  $66$  out to  $318 R_\odot$  [Howard *et al.*, 2008].

Even in the earliest coronagraph images from spacecraft, it was noted that CMEs can change trajectory after being initiated [Liewer *et al.*, 2015]. Recently, these deflections have been associated with the balance of magnetic pressure forces on the CME and consequently the location of the heliospheric current sheet (HCS) [Liewer *et al.*, 2015]. The HCS forms in the solar corona when plasma flows with opposite magnetic polarity meet, creating a region of reduced magnetic pressure around discontinuity bounding the different polarities [Smith, 2001].

Shen *et al.* [2011] studied the deflection of one CME in detail using COR1, COR2, and EUVI images from the SECCHI instrument package on the STB spacecraft. Using a model to simulate the magnetic energy density in the solar corona, they found that the CME was deflected toward the HCS. Gui *et al.* [2011] studied this further by considering 10 different CMEs in data from the STEREO spacecraft and found that eight of these were also deflected toward the HCS. Liewer *et al.* [2015] analyzed five CMEs using data from both STEREO spacecraft, allowing the complete three-dimensional trajectory to be observed to  $15 R_\odot$ . The CMEs were chosen to be well defined, with only one CME occurring in the images at once. Four of the CMEs showed large deflections of  $25$ – $40^\circ$  in latitude and  $10$ – $25^\circ$  in longitude. The final CME was not deflected and was chosen as a control. A Potential Field Source Surface (PFSS) model was used to estimate the direction of the magnetic pressure force and the location of the HCS. PFSS models have a good resolution and are relatively simple to implement, but they do not consider time-dependent phenomena Riley *et al.* [2006]. Liewer *et al.* [2015] found that the CMEs deflected to the HCS but that more local solar features also caused deflection. Finally, Kay *et al.* [2015] have developed Forecasting a CME’s Altered Trajectory (ForeCAT), a model which forecasts the deflection of CMEs based on input parameters including the shape and location of the CME. Using a PFSS model to estimate the solar magnetic field, ForeCAT was tested on CMEs in April and May 2005. ForeCAT predicted that most of the CMEs would deflect to the HCS, with weaker deflections during solar minimum.

The remote sensing data returned by the SECCHI package must be processed to identify and characterize CMEs. A catalogue of CMEs derived from the SECCHI COR2 and HI data by expert identification has recently been released as part of the Heliospheric Cataloguing, Analysis and Techniques Service (HELCASTS) project (<https://www.helcats-fp7.eu/>). Currently, a catalogue of CMEs found in Large Angle and Spectrometric Coronagraph (LASCO) images by manual identification is available, the Coordinated Data Analysis Workshop (CDAW) catalogue, created from a workshop in 1999 [Gopalswamy *et al.*, 2009]. However, the method of expert identification of CMEs is not without its limitations. A CME study by Webb and Howard [2012] highlight that the CDAW catalogue has had at least four different experts identifying the CMEs in its lifetime, which introduces biases as different experts might have different views on what counts as a CME. Another issue with manually examining images is that it is time consuming and subjective [Yashiro *et al.*, 2008].

Limitations of manual identification, such as those mentioned above, motivated the development of automated algorithms to identify CMEs. These have several advantages, for example, the results are repeatable [Yashiro *et al.*, 2008]. Examples of these include the Solar Eruptive Event Detection System (SEEDS), Computer Aided CME Tracking Software (CACTus), and Coronal Image Processing (CORIMP). SEEDS is based on a two-dimensional to one-dimensional projection method which defines CMEs as bright spots with increased density moving radially outward [Olmedo *et al.*, 2008]. CACTus identifies CMEs by using the Hough transform to find bright points in time and height [Robbrecht and Berghams, 2004]. CORIMP detects CMEs by splitting coronagraph images into two parts; the first part is the quiescent part containing the background corona

and features such as streamers that do not change quickly in time, the second is the dynamic part containing features that change rapidly in time, namely, CMEs [Morgan *et al.*, 2012]. Both the SEEDS and CACTus software have been applied to SECCHI COR2 coronagraph images.

More recently, another option has been explored: the Solar Stormwatch (SSW) project, a Zooniverse citizen science project jointly developed by the Rutherford Appleton Laboratory and the Royal Observatory Greenwich. Zooniverse is a platform for hosting citizen science projects, connecting researchers with volunteers, which has run many successful projects such as Galaxy Zoo, Planet Hunters, and Old Weather. The Solar Stormwatch project combines the observations of many citizen scientists to identify CMEs in SECCHI images. This method allows the quantitative estimate of the uncertainty in computed CME properties, which can be found from the distributions of estimates provided by the great number of volunteers taking part. Data from the Solar Stormwatch project have already been used to create a catalogue of CMEs from the SECCHI HI imagers on the STEREO spacecraft [Barnard *et al.*, 2014, 2015b]. The Solar Stormwatch approach is described in section 2.1.

This work analyzes the evolution of a set of CMEs through the STEREO EUVI, COR1, and COR2 fields of view, by producing a new catalogue of events derived from the citizen science estimates of the CME timing, location, and size provided by the Solar Stormwatch Track-it-back activity. We assess how the CME speeds, directions, and widths vary as the CMEs propagate through the fields of view of the different imagers. In particular, we compare the observed CME deflections between the COR1 and COR2 fields of view to the separation between the CME source region and the HCS. Section 2 of this paper details the methods used to process the Solar Stormwatch identifications and create the list of CMEs matched between the COR1, COR2, and EUVI data. Section 3 presents the results of analysis of this event list, in particular, a study of deflections of the CME trajectories. Finally, the results are discussed in section 4.

## 2. Methodology

### 2.1. Solar Stormwatch and Track-It-Back

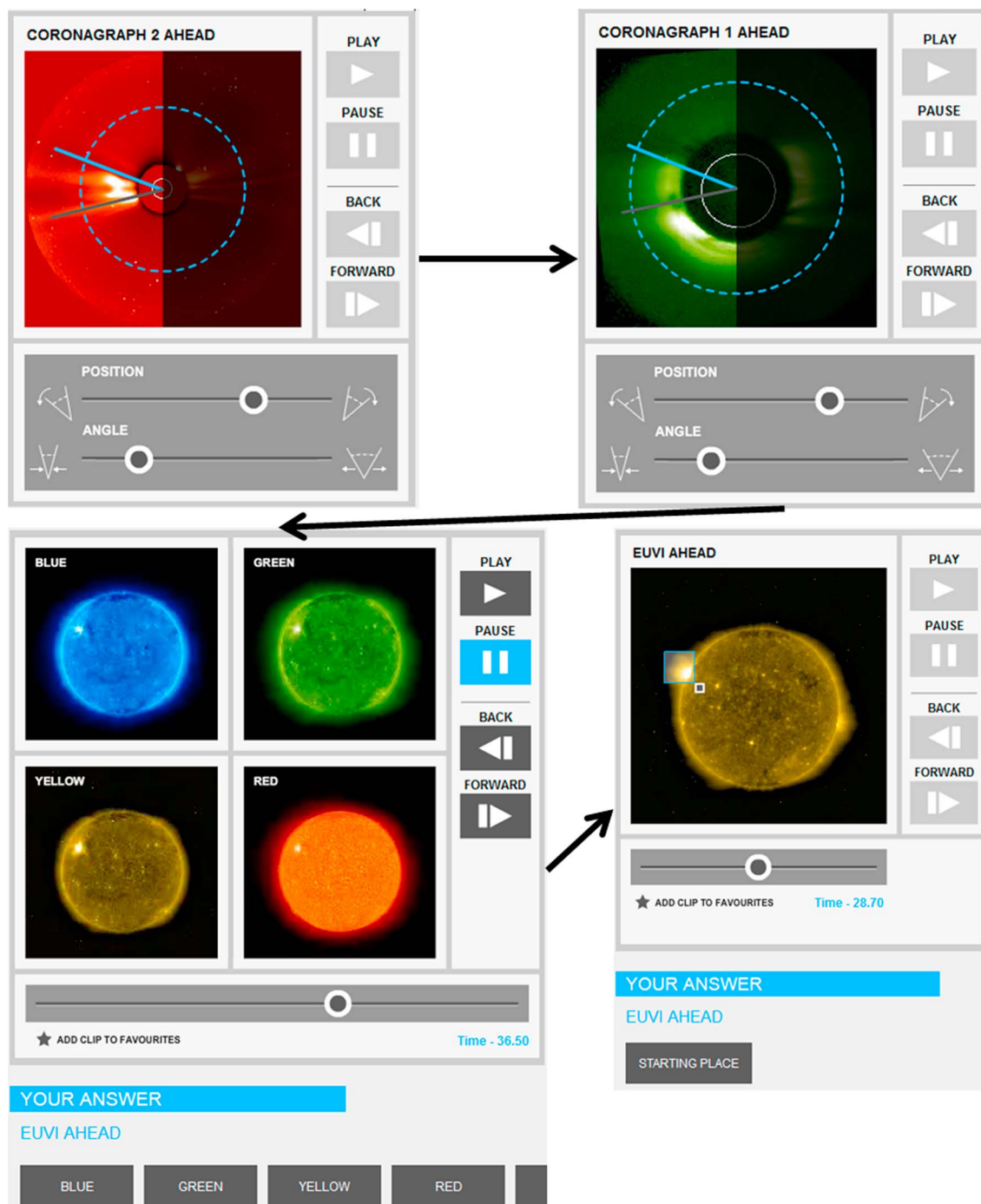
People who participated in the Solar Stormwatch project were asked to complete various activities involving the identification and classification of CMEs in SECCHI images. These included the following:

1. "Spot!" and "Trace-it!" involved identifying CMEs in images from HI1 and tracking their propagation through the inner solar system, respectively.
2. "Incoming!" and "Incoming Trace-it!" were similar to the two just described but used STEREO beacon data (HI images transmitted in real time at a lower resolution).
3. In "What's that?" participants were asked to mark anything unusual in the HI1 and HI2 images, such as a comet or dust impact.
4. "Track-it-back" asked participants to classify CMEs in COR1, COR2, and EUVI SECCHI images.

A detailed summary of these activities is provided by Barnard *et al.* [2014], and a review of Solar Stormwatch investigations is given by Barnard *et al.* [2015a].

However, in this paper we focus on the results of the Track-it-back exercise, which involved tracking CMEs found in the HI1 and HI2 data back through the COR2, COR1, and EUVI images. An example of this exercise is illustrated in Figure 1. Participants were asked to mark the first appearance time of the CME in the coronagraph images; the time the CMEs reached halfway through the images (hereafter referred to as the "midpoint time," shown by the blue dashed lines in Figure 1); the position angle of the top edge of the CME (the angle from solar north at which the CME occurred, shown as the solid blue line in Figure 1); and the apparent angular width (shown as the difference between the blue and grey lines in Figure 1). Finally, participants had to choose which of the four EUVI wavelength images showed the CME most clearly and mark the appearance time of the CME and the pixel coordinates of the estimated source region of the CME in the image (shown in the lower right box of Figure 1).

The images analyzed in this exercise were from 28 February 2007 to 12 February 2010, as these were the data available when the project was created. The images were all rescaled to  $256 \times 256$  pixels, and the movies were formed from frames with a cadence of 15 min. The COR2 images were intensity plus polarized images, and the COR1 images were polarized images. The analysis was limited by restrictions in average household internet bandwidth at the time the project was instigated. Improvements could now be made with faster internet speeds and advances in image processing techniques [Druckmullerova *et al.*, 2011], [DeForest *et al.*, 2011].



**Figure 1.** The various stages of the Solar Stormwatch project Track-it-back activity. (top left) Participants first marked the appearance time, the time the CME reached the blue circle, and the position angle and width of the CME in the COR2 field of view. (top right) Participants marked the appearance time, the time the CME reached the blue circle, and the position angle and width of the CME in the COR1 field of view. (bottom left) Participants chose which EUVI wavelength showed the CME most clearly. (bottom right) Participants identified when the CME appeared in the EUVI field of view and drew a box around the CME source region.



**Table 1.** The Total Number of Submissions in the COR1 and COR2 Data Sets, Showing the Number of Submissions That Failed for Different Reasons

Number of Submissions	COR1	COR2
Initial submissions	8,015	10,548
Position angle = 360°	1,227	731
Width = 0	1,050	630
Time taken < 0	2,140	2,160
Speed < 50 km/s	2,236	2,148
Speed > 3000 km/s	856	1,149
Total removed	3,360	3,552
New total submissions	4,655	6,996

This Track-it-back exercise resulted in three data sets: identifications in the COR1, COR2, and EUVI images; these were composed of 8015, 10,548, and 5238 classifications, respectively. These three data sets combined provide the opportunity to study the evolution of CMEs through the whole field of view spanned by EUVI-COR1-COR2, such as the acceleration through the corona and deflection of the CME trajectory.

## 2.2. Quality Control

First, the data had to be checked for quality and consistency. This was done by removing any COR1 or COR2 observations for which the CME width or position angle were reported

to be 0 or 360°, respectively, as these were the default values returned when a participant submitted a classification without tracking a CME. Furthermore, classifications were excluded if the difference between the appearance and midpoint times was less than or equal to 0, as this would imply that the CME was propagating toward the Sun.

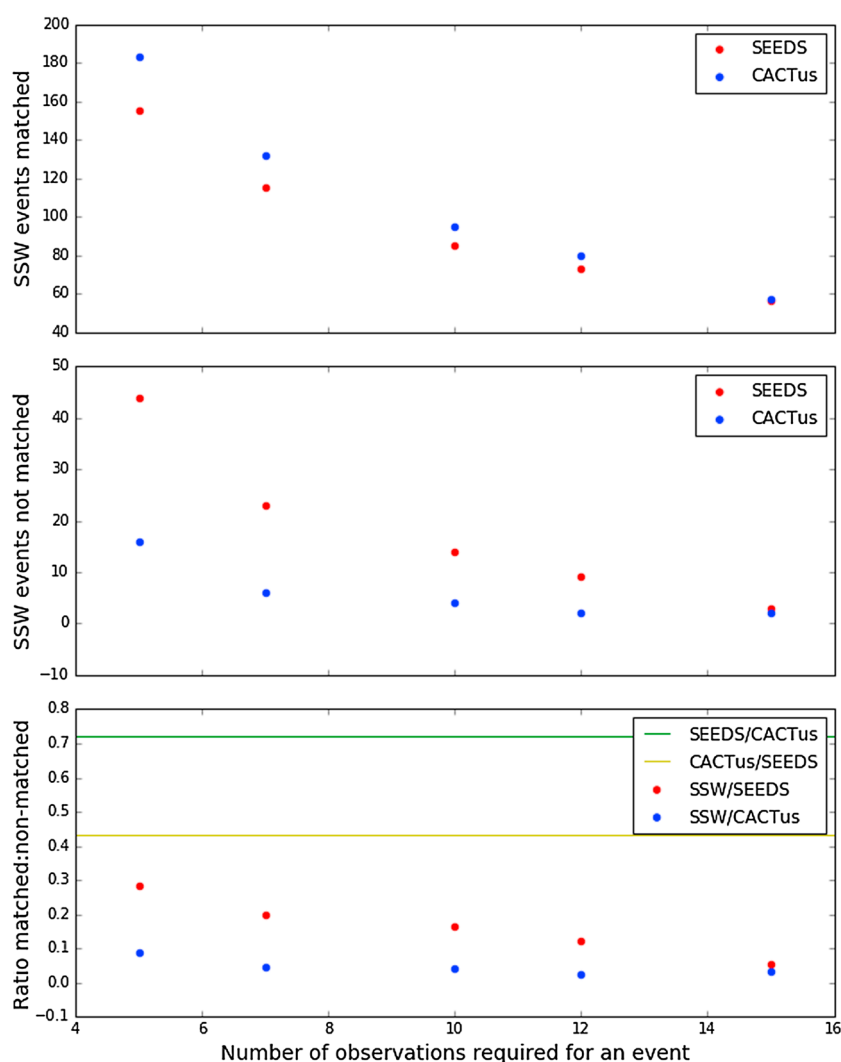
The differences between the appearance and midpoint times were used to calculate plane-of-sky (POS) speeds for both the COR1 and COR2 data. The mean distances that the CME had to travel between those times were found by looking at images from all months of the year and averaging the POS distances at the middle point of the field of view and at the edge of the occulting disk, which defined the appearance distance. This was necessary as the spacecraft each have different orbits which are elliptical, causing these distances to vary a small amount throughout the orbit. These changes were less than 1.5% of the mean distance throughout the orbit. The mean distances for COR1 were 1.22 and 1.32  $R_{\odot}$  for STA and STB, respectively, and the mean distances for COR2 were 6.31 and 7.00  $R_{\odot}$  for STA and STB, respectively. The speeds were then calculated by dividing the appropriate distances by the corresponding times to obtain COR1 and COR2 speeds. Using the distribution of CME speeds from *Gopalswamy et al.* [2009] as a guide, a criterion was devised to ensure that each CME classification had a realistic speed. Therefore, any classifications that produced speeds below 10 km/s or above 3000 km/s in the COR1 data, or below 50 km/s or above 3000 km/s in the COR2 data, were removed from the data sets.

Table 1 shows the number of submissions in the data set that were removed for the reasons stated above. The total number of submissions removed is not equal to the sum over each rejection criterion as many submissions failed for multiple reasons. In total, 58 % of the COR1 and 66 % of the COR2 data were selected for analysis.

## 2.3. Clustering

The clustering method used in this analysis is a modified version of the methods used in *Barnard et al.* [2014] and *Tucker-Hood et al.* [2015]. Observations were split into Carrington rotation blocks, and the clustering procedure was applied to each block separately. The position angle of the top edge and angular width were used to find the central position angle of each classification, by adding half the width to the position angle of the top edge. The observations for the block were then placed into a time and position angle grid, with a resolution of 1 h and 2°, using the midpoint times and central position angles. The number of observations at each point in the grid was then counted, using a rolling position angle window of  $\pm 15^\circ$  and rolling time window of  $\pm 1$  h. Regions of the position angle time grid which included more than 12 observations in the COR2 data, or 7 observations in the COR1 data, were identified as a cluster that likely represents the identification of a CME.

The thresholds of  $\pm 15^\circ$ ,  $\pm 1$  h, and 12 and 7 observations were deduced after experimenting to find a set of values that were judged to work reasonably well. The clustering method described above was run multiple times using different thresholds; using time windows of 1 to 2 h; position angle windows of 10 to 25°; and minimum number of observations ranging from 5 to 20. The event list from each of these runs was then matched to events in the SEEDS and CACTus catalogues using an algorithm that will be described in section 2.5. Any changes between the event lists were then explored.



**Figure 2.** The investigation of COR2 threshold number of observations. (top) The total number of SSW events matched to the SEEDS and CACTus catalogues. (middle) The number of SSW events not matched to the SEEDS and CACTus catalogues. (bottom) The ratio of nonmatched events to matched events, where the green line is the ratio of events in the SEEDS catalogue not matched to events in the CACTus catalogue and the yellow line is the ratio of events in the CACTus catalogue not matched to events in the SEEDS catalogue.

The position angle and time windows were found to be of little importance; the number of events varied very little when these were changed, so these were chosen to be 1 h and 15°. However, the number of identified events was found to be very sensitive to the threshold number of observations included in each CME cluster. Five different values were tried for this threshold: 5, 7, 10, 12, and 15. Figure 2 compares how the number of events identified by SSW could be matched, or not, to events from the SEEDS and CACTus catalogues, as a function of the number of SSW classifications required to define an event. Figure 2 (top) shows the number of events that could be matched, Figure 2 (middle) shows the number of events that could not be matched, and Figure 2 (bottom) shows the ratio of these numbers. Also, included in Figure 2 (bottom) is the ratio of SEEDS events that could be matched to CACTus events (green line) and the ratio of CACTus events that were matched with SEEDS events (yellow line). It was found that the ratio of nonmatched to matched events increased rapidly as the threshold was reduced, which is shown in Figure 2. The threshold value of 12 was chosen so that this ratio was low when comparing with both the SEEDS and CACTus catalogues. Various thresholds were then applied for the COR1 data, keeping the time and position angle windows the same as in COR2. Through these tests, a threshold of 7 was chosen for the COR1 data, which struck a balance between being sensitive to poorly defined events and extracting false positives.



For each cluster of CME characterizations in a Carrington rotation, the individual characterizations were extracted and grouped together for that event. This method was repeated for each Carrington rotation block. Figures showing the time and position angle grid with identified clusters are available online (COR1: <https://doi.org/10.6084/m9.figshare.4747927.v1>, COR2: <https://doi.org/10.6084/m9.figshare.4747936.v1>). For each CME we had information from COR1 and COR2 on appearance time, midpoint time, central position angle, angular width, and number of observations combined to calculate the properties of the CME. Each property was calculated as the mean of all values from the identifications within each cluster. This was appropriate as examining the classifications from several example events demonstrated that the classifications of each quantity were approximately normally distributed. The errors for each of these properties were calculated as the standard error of the mean. This method was applied to both the COR1 and COR2 data sets and identified 86 and 82 CMEs, respectively.

To create a catalogue of CMEs containing information from both the COR1 and COR2 data sets, the COR1 and COR2 CME lists were compared and a new list compiled of the events which appeared in both of these lists. This was achieved by finding CME events from the COR1 and COR2 lists which occurred within 12 h of each other and within a position angle of  $45^\circ$ . A criterion was also applied to ensure that the COR1 midpoint time was before the COR2 midpoint time, as CMEs pass through the COR1 frame first. From the events which fitted this criteria, the closest CME in time in the COR1 list was matched to each event in the COR2 list. In total, 41 CMEs were matched between the COR1 and COR2 events lists, 20 of which were from STA images and 21 from STB. This matched event list uses 22% and 23% of the total submissions from the COR1 and COR2 data sets.

To add the EUVI data to the COR1 and COR2 event list, all the EUVI observations which listed an appearance time in the 12 h before the COR1 start time were combined, and the parameters were calculated as the mean of the observations, and errors as the standard error. Including EUVI observations with a large window size of 12 h before the COR1 appearance time was necessary to account for the large uncertainties of the COR1 appearance time, which will be discussed in section 3.1.1. Unfortunately, there were no matching EUVI classifications for four events shown in STA images and three events in STB images. The EUVI parameters were the appearance time, the pixel coordinates of the size and location of the box the participant drew over the source of the CME, the number of observations used to calculate the EUVI parameters, and the EUVI wavelength identified as most clearly showing the eruption. As the number of EUVI classifications was much lower than the number of COR1 and COR2 classifications, the CME source regions and EUVI appearance times were manually checked in the EUVI images. For 22 of the SSW events, the EUVI source location and appearance times matched up to activity in the images, while for the remaining events the CME source regions matched up to the locations of active regions on the Sun, though we could not be certain whether the CMEs originated from these regions.

#### 2.4. Additional COR1 and COR2 Parameters

The properties of the CMEs in the event list described above were used to calculate several additional parameters. As previously stated, the STEREO spacecraft each have slightly different elliptical orbits which vary by less than 1.5% of the mean distance throughout the orbit. In the quality control section 2.2, speeds were calculated using the mean POS distances from the edge of the occulting disk to the middle point of the field of view throughout the year. However, to improve the accuracy of these distance calculations, the COR1 and COR2 data were analyzed for each CME, to compute the average POS appearance and midpoint distances over the duration of the event. COR1 and COR2 speeds were then calculated by dividing the POS distance traveled between the appearance and midpoint locations by the time difference between the CME appearance and midpoint times. The errors in the appearance and midpoint times were assumed to be uncorrelated and were added in quadrature to give an estimate of the errors in the speeds [Harrison, 2015].

POS latitudes were estimated from the position angle data, as these allow the STA and STB data to be analyzed more easily on the same axis. For position angles less than  $180^\circ$ , latitude was calculated as  $90^\circ$  minus the position angle. For position angles greater than  $180^\circ$ , latitude was calculated as the position angle minus  $270^\circ$ . The procedure to convert position angles into latitudes is an approximation which does introduce a variable error. This systematic error depends on how close the CME is to the plane of the sky, how close the CME is to the equator and the solar B0 angle (the tilt of the Sun's rotational axis compared to the solar equatorial plane), and overestimates the true latitudinal position, as highlighted by Liewer *et al.* [2015]. This error is minimized for CMEs propagating near the equator and close to the plane of the sky, for small B0 angles.

POS deflections between the COR1 and COR2 fields of view were calculated as the POS latitude in COR2 minus the POS latitude in COR1. The errors in the latitudes were also assumed to be uncorrelated and were added in quadrature to provide uncertainty estimates for the CME deflections [Harrison, 2015].

Finally, the approximate source locations of the CMEs on the solar disk were estimated to be the average of the centroids of the boxes the participants had drawn on the EUVI images. This was checked by downloading EUVI images of each event from the Virtual Solar Observatory (VSO) using the SunPy Python package [SunPy Community *et al.*, 2015] and superimposing the locations of the source locations onto the images. In four cases the midpoint of the box was not on the solar disk, so an approximate location on the solar disk was estimated by radially projecting the point back toward the solar limb in the plane of the sky. The pixel coordinates were then converted into Carrington coordinates. A possible explanation for the locations of these four source regions might be that the boxes the citizen scientists drew were not centered on the source region, meaning that the centroids of these boxes were not actually over the source region. It is also possible that the actual source region was on the solar limb, or on the far side of the disk, and the citizen scientists were just identifying where this material appeared in the EUVI field of view.

### 2.5. Matching Events to SEEDS and CACTus

To evaluate how well the citizen scientists isolated CME parameters, the Solar Stormwatch CMEs were matched to CMEs in the SEEDS and CACTus catalogues. As the SEEDS and CACTus software have only been applied to the COR2 data, this section refers to the COR2 SSW data only.

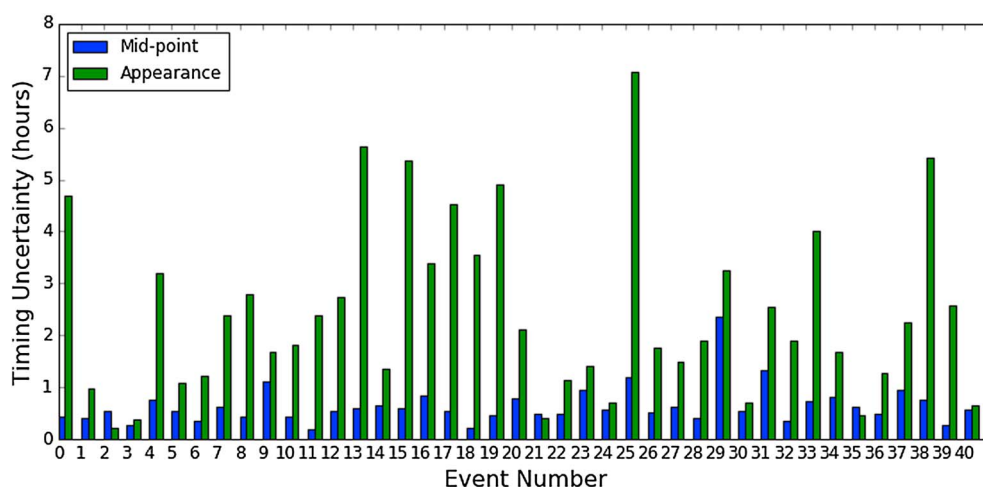
An algorithm was created to match the SSW events to the SEEDS and CACTus events, which worked by finding the differences between the SSW midpoint times and the SEEDS or CACTus start times. The SSW midpoint time was used rather than the appearance time due to the larger errors in appearance time estimates compared to the midpoint time; this is described in section 3.1.1. The SEEDS or CACTus events closest to the SSW events in time were matched together, ensuring that the SEEDS and CACTus position angles were within  $45^\circ$  of the SSW position angle. This algorithm was also used to determine the thresholds for clustering classifications of CMEs as described in section 2.3.

Once the above algorithm had been used to match the SSW events to the SEEDS and CACTus events, an extra check was performed on the matched CMEs to ensure that they were the same events as the SSW CMEs. The differences between the SSW event midpoint time and matched SEEDS or CACTus event times were calculated, and any matched events that had a start time which differed by more than 12 h to the SSW time were flagged. Two of the CACTus matches did not meet this time criterion; these SSW events had no viable matches in the CACTus catalogue and had been matched to different events. One of these two events could also not be matched to an event in the SEEDS catalogue. Manual inspection of the coronagraph images show that a CME did occur in both cases. However, in each case the CME appeared to be released at the location of a streamer, possibly explaining why the CACTus and SEEDS software failed to identify the CME.

### 2.6. Comparing CME Deflections to CME Separation From the HCS

Finally, deflections in the trajectories of the CMEs were investigated by comparing the latitudinal separation of the CME source location from the HCS against the POS deflections. A description of how the source locations of the CMEs were found in Carrington coordinates, and a description of how the POS deflections were calculated were given in section 2.4. The location of the HCS at  $8.75 R_S$  was estimated from the output of the MAS MHD model of the solar corona [Linker *et al.*, 1999; Riley *et al.*, 2011]. Simulations of each Carrington rotation were used to compute an estimate of the heliographic latitude of the HCS at the same heliographic longitude of the CME source region. The HCS location at  $8.75 R_S$  was chosen as this is the approximate distance at which the CMEs pass through the radial midpoint of the COR2 images. The locations of the HCS were then compared to the source locations of the CMEs in Carrington coordinates to give estimates of the latitudinal separation between the two. To give an estimate of the errors on these separations, the latitudinal separation from the CME source locations to the HCS were also found for events  $\pm 10^\circ$  latitude, and the errors were taken to be the magnitude of the differences in latitudinal distance to the HCS.

To investigate the association between the latitudinal distances and the POS deflections, a linear regression technique was applied. This was implemented using the linear regression function in the SciPy Python package (scipy.stats.linregress), which calculates a linear least squares regression, returning the slope, intercept, and standard error of the regression line. The function calculates a two-sided *t* test [Wilks, 2011] to test the null hypothesis that the slope is 0 against the alternative hypothesis that the slope is not 0, and the slope and



**Figure 3.** Magnitudes of uncertainties (in hours) on the COR2 appearance times (green) and midpoint times (blue).

intercept of the regression line were then used to calculate the predicted  $y$  values of the linear fit, by using the formula:  $y = \text{intercept} + \text{slope} \times x$ .

The deflections observed in coronagraph images are not always representative of the true latitudinal deflection. This is because projection effects as well as longitudinal deflections of the CME can change the observed latitudinal deflection. To account for this issue, limb events (events with source regions at the edge of the solar disk, as seen in the coronagraph images) were considered separately. The POS latitudinal deflections of these events are likely to be a better approximation for the true latitudinal deflection. Therefore, the linear regression technique described above was repeated for limb events only. The EUVI images created to find the CME source locations as described in section 2.4 were used to assess which events were limb events.

### 3. Results

#### 3.1. Properties

##### 3.1.1. Timings

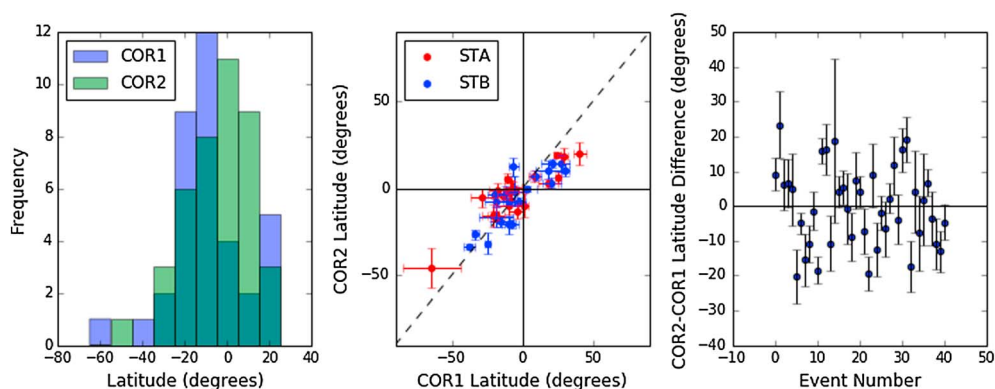
The SSW event list contained 41 CMEs in the time period from May 2007 to February 2010. The CMEs did not occur uniformly over this period, however; over half the CMEs occurred within the last few months of the investigated period. The lower CME occurrence rate over the period 2007–2009 was due to the deep minimum in solar activity at the end of solar cycle 23. Solar activity, and hence the CME rate, started to rise again in 2010, on the ascending phase of solar cycle 24. This is consistent with the findings of *Gopalswamy et al.* [2009], *Robbrecht et al.* [2009], and *Barnard et al.* [2014].

The catalogue included four different times for each CME: the time the CME appeared in the COR1 and COR2 fields of view and the time it reached the radial midpoint of the images. Figure 3 shows the errors in the COR2 midpoint and appearance times. The appearance time errors are shown in green, and the midpoint time uncertainties are shown in blue. This shows that there was much variability in the timing uncertainties between events. On average, the appearance time uncertainties are 4 times larger than the midpoint time uncertainties, and only for two events is the appearance time uncertainty less than the midpoint time uncertainty.

##### 3.1.2. Latitudes

POS latitudes were calculated from the position angles for both the COR1 and COR2 data sets, as described in section 2.4. Figure 4 (left) is a histogram of the COR2 latitudes, created using a bin with of  $10^\circ$ , and shows that the CMEs in the SSW catalogue are clustered around the solar equator.

Figure 4 (middle) shows the COR1 latitudes plotted against the COR2 latitudes for each CME. The scatter of the CME POS latitudes around the one-to-one line shows that the CMEs tended to change latitude between the COR1 and COR2 fields of view, implying that CMEs are typically deflected between the two fields of view. This is also demonstrated in Figure 4 (right) which shows the COR2 minus COR1 latitude differences, where many of the events are far from the zero line. In general, the direction of the deflections appears to be toward the



**Figure 4.** (left) Histogram showing the distribution of COR1 and COR2 latitudes. (middle) Scatterplot showing the change in latitude observed between the COR1 and COR2 fields of view. Events observed by STA are shown in red, and events observed by STB are shown in blue. (right) Differences between the COR2 and COR1 latitudes for each CME.

equatorial region. The mean deflection size was  $10^\circ$ , though the deflections varied from  $1$  to  $23^\circ$ . The errors of the SSW deflections were between  $3$  and  $23^\circ$ .

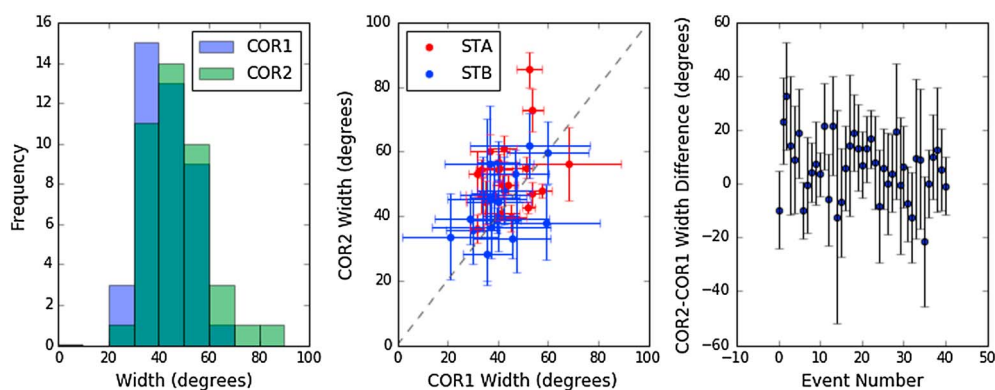
### 3.1.3. Apparent Widths

The event list contained the apparent angular widths found in both the COR1 and COR2 images; Figure 5 (left) is a histogram of the COR2 angular widths using a bin width of  $10^\circ$ . This shows that most of the CMEs had a width of between  $30$  and  $60^\circ$  with the mode of the distribution occurring at  $40^\circ$ .

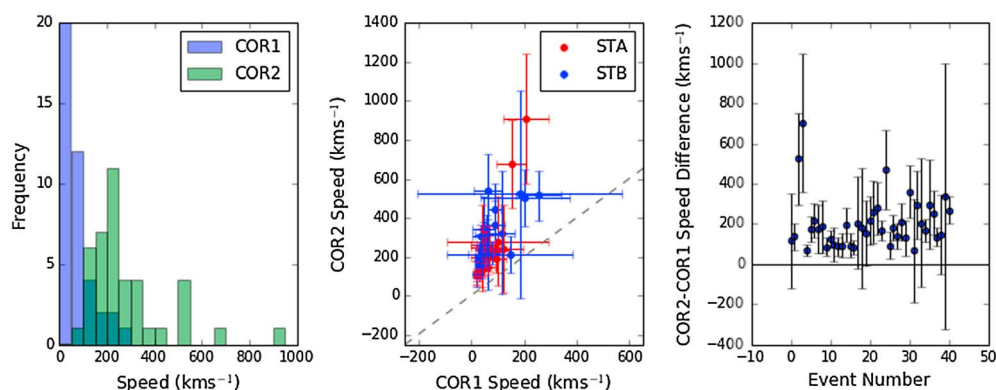
Figure 5 (middle) shows the COR1 apparent widths plotted against the COR2 apparent widths, and Figure 5 (right) shows the differences between the COR1 and COR2 width estimates. The mean absolute difference in widths was  $11^\circ$ . Of the 41 width differences shown, 28 are above the zero line, suggesting that CMEs tend to appear to increase in width from COR1 to COR2. Although this growth could be attributed to a possible lower image quality in the COR1 instrument compared with the COR2 instrument, the uncertainty in the CME width estimates from the COR1 and COR2 data were similar in magnitude, suggesting that participants did not find it more difficult to classify CME widths in the COR1 images. Therefore, we consider that it is likely that this result does imply a growth in CME width between the COR1 and COR2 fields of view. The other 13 are below the line, but all the errors on these points cross the zero line; therefore, we do not have any robust observations of a CME appearing to decrease in width between the COR1 and COR2 fields of view.

### 3.1.4. Speeds

Figure 6 (left) is a histogram of the COR2 speeds created using a bin width of  $50$  km/s, which shows that the majority of the CMEs in the SSW catalogue were slow CMEs, with  $908$  km/s the fastest CME speed found. The mean COR1 and COR2 speeds were  $75$  km/s and  $280$  km/s, respectively.



**Figure 5.** (left) Histogram showing the distribution of COR1 and COR2 widths. (middle) Scatterplot showing the change in widths observed between the COR1 and COR2 fields of view. Events observed by STA are shown in red, and events observed by STB are shown in blue. (right) Differences between the COR2 and COR1 widths for each CME.



**Figure 6.** (left) Histogram showing the distribution of COR1 and COR2 speeds. (middle) Scatterplot showing the change in speed observed between the COR1 and COR2 fields of view. Events observed by STA are shown in red, and events observed by STB are shown in blue. (right) Differences between the COR2 and COR1 speeds for each CME.

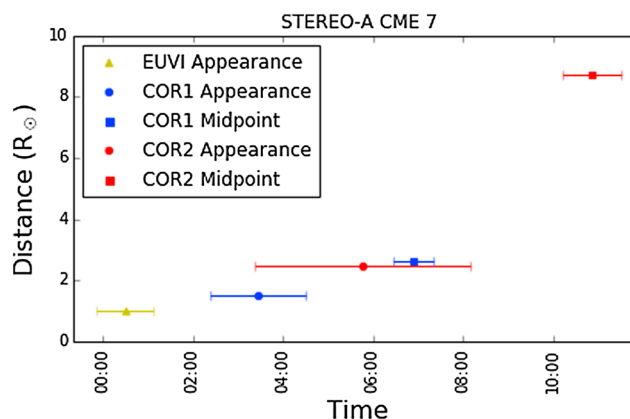
Figure 6 (middle) shows the COR1 speeds plotted against the COR2 speeds. The COR1 speeds were subtracted from the COR2 speeds to find the speed differences between the two fields of view. Figure 6 (right) displays these differences and shows that the CMEs all accelerated between COR1 and COR2; the mean acceleration was  $204 \pm 127$  km/s. Of the 41 events shown in this plot, 33 have error bars which are above the zero line; therefore, in general, it is unlikely that these positive acceleration estimates are due to uncertainties in the speed estimates.

Finally, time versus distance plots for each event showed the acceleration of each CME; the full set of plots is available online <https://doi.org/10.6084/m9.figshare.4725874.v1>. Figure 7 shows the plot for one CME which was seen by STA. This illustrates just how large the appearance time errors are compared to the midpoint time errors; the COR2 appearance time is before the COR1 midpoint time. Generally, these plots indicate that the CME speed evolves nonlinearly through the combined EUVI, COR1, and COR2 FOV. This is a potential limitation in the analysis of speeds presented in this section.

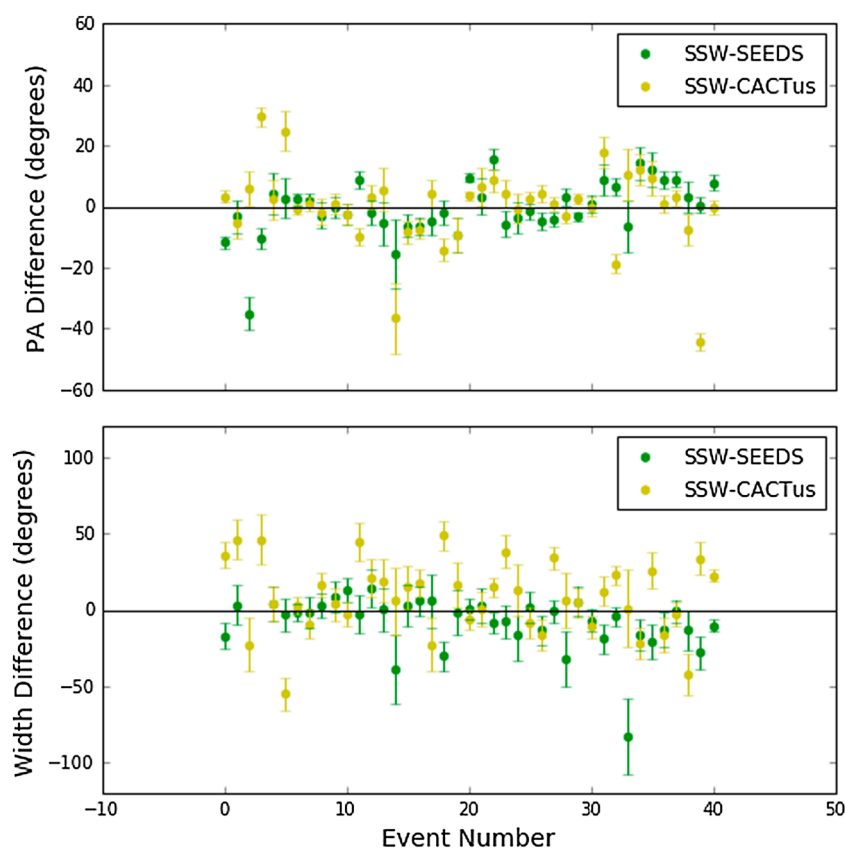
### 3.2. Comparisons

This section describes the results of comparing the SSW event list with the SEEDS and CACTus catalogues. As the SEEDS and CACTus software have only been applied to the SSW COR2 data, this section only considers the COR2 and not the COR1 data. Here we compare the CME timings, position angles, widths, and speeds.

Figure 8 (top) shows the SSW-SEEDS and SSW-CACTus position angle differences. The mean SSW-SEEDS position angle difference was  $-1^\circ$ , and the mean SSW-CACTus position angle difference was  $0^\circ$ , though the SSW position angle errors ranged from  $1$  to  $11^\circ$ . Fair agreement was to be expected as the position angles were used to match the SSW events to events in the SEEDS and CACTus catalogues by ensuring the differences were below  $45^\circ$ ; however, this agreement is much closer than the matching criterion.



**Figure 7.** Time versus distance plot for one event.



**Figure 8.** (top) Differences between the SSW and SEEDS position angles (green) and SSW and CACTus position angles (yellow). (bottom) Differences between the SSW and SEEDS widths (green) and SSW and CACTus widths (yellow). The SSW-SEEDS width differences for events 2 and 3 are not shown as they are  $-170^\circ$  and  $-291^\circ$ , respectively, and would therefore make the plot difficult to interpret. The SSW-SEEDS width difference appears to be missing for event 4, but this is actually because it is the same value as the SSW-CACTus difference. Errors shown for both plots are the SSW errors as SEEDS and CACTus catalogues do not publish error estimates.

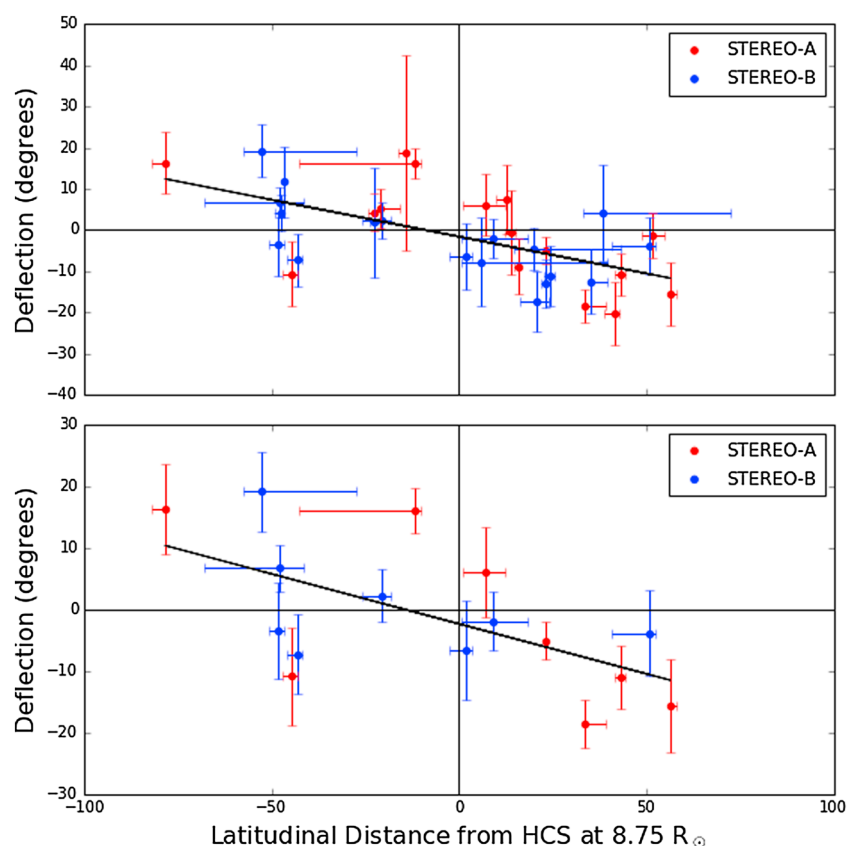
Figure 8 (bottom) shows the SSW-SEEDS and SSW-CACTus apparent width differences. The mean SSW-SEEDS width difference was  $-19^\circ$ , and the mean absolute difference was  $22^\circ$ . The mean SSW-CACTus width difference was  $8^\circ$ , but the mean absolute difference was  $20^\circ$ . However, the SSW width errors were between  $4^\circ$  and  $25^\circ$ , so the differences could have been due to these errors.

The mean SEEDS-CACTus, SSW-SEEDS, and SSW-CACTus speed differences were  $+354$ ,  $-397$ , and  $-43$  km/s, respectively. The mean absolute SEEDS-CACTus, SSW-SEEDS, and SSW-CACTus speed differences were  $477$ ,  $444$ , and  $136$  km/s, respectively. Finally, the median SEEDS-CACTus, SSW-SEEDS, and SSW-CACTus speed differences were  $9$ ,  $-43$ , and  $-41$  km/s. The median values were considered as the mean differences were likely skewed by the few events for which the speed differences were comparatively large. On average, the SSW speeds are slower than both the SEEDS and CACTus speeds. This is likely a consequence of the fact that the SSW appearance times were earlier than the CACTus and SEEDS times. However, the SSW speeds agreed with the CACTus speeds well; the SEEDS speeds were much faster than both the SSW and CACTus speeds. Note that the SSW speed errors ranged from  $27$  km/s to  $533$  km/s.

### 3.3. Deflections

Figure 4 (right) shows the POS latitude deflections of the CMEs between the COR1 and COR2 fields of view. POS latitude deflections were calculated and compared to latitudinal distances between the source region of the CME and the HCS found using the MAS model as described in section 2.6. Figure 9 (top) shows the POS latitude deflections plotted against the latitudinal distances from the HCS for the 34 CMEs with EUVI data. The plot shows that all but six CMEs deflected toward the HCS, though the error bars of four of these are large enough that this could have been due to the uncertainties in the data. A linear least squares regression was applied to the data and is shown as a black line in the plot. The slope of this line was found to be significantly





**Figure 9.** POS latitude deflections plotted against latitudinal distance from the heliospheric current sheet (HCS). STA events are shown in red, STB events are shown in blue. (top) All 34 deflections and (bottom) only limb events.

different from zero, with a  $p$  value of  $1.6 \times 10^{-4}$  using the two tailed hypothesis test discussed in section 2.6. There appears to be an approximately linear relationship between the latitudinal separation of the CME source region and HCS and the magnitude of the CME deflection toward the HCS.

Figure 9 (bottom) shows the POS latitude deflections against the latitudinal distance from the HCS for the subset of CMEs identified as limb events. This plot shows a similar trend between deflections and the latitudinal distance to the HCS. A linear least squares regression was also applied to this data, yielding a similar relationship to that obtained when all events were considered. The  $p$  value for this test was  $1.4 \times 10^{-2}$ .

#### 4. Discussion and Conclusions

This paper explored a data set from the Solar Stormwatch Track-it-back activity, covering the time period from 28 February 2007 to 12 February 2010. The COR1 and COR2 submissions were clustered into events, and a matched catalogue of 41 CMEs was created, which contained 20 CMEs observed by STA and 21 CMEs observed by STB. The angular separation of the spacecraft increases from  $1^\circ$  at the start to  $136^\circ$  by the end of the time period; therefore, there were times when the spacecraft were observing similar parts of the Sun. A preliminary look at the data suggests that several of the events in the catalogue could actually be the same events observed by both STA and STB, meaning that the SSW event list might contain fewer than 41 unique events. This matched event list only used 22% and 23% of the total COR1 and COR2 submissions, respectively; therefore, there was a substantial amount of data that was not used.

The number of observations varied dramatically for the different imagers; there were 10,548 COR2 submissions and 8015 COR1 submissions but only 5238 EUVI submissions. This meant that there were comparatively fewer EUVI observations to match to the CMEs identified in the COR1 and COR2 data, which is a limitation to the conclusions drawn in this paper. As the Track-it-back exercise asked participants to identify CMEs first in the COR2 images, then COR1 and then EUVI, it would be worth investigating the reasons for this, for example; did the participants find the exercise too difficult?

The citizen science approach was found to be effective at isolating the time the CMEs reach the midpoint of the POS in the coronagraph images; however, they were less successful in isolating the appearance time of the CME in the COR1 and COR2 images. This was indicated by the larger uncertainties on the appearance times, being on average 4 times larger than the midpoint times. In a future study, this issue could be solved by asking participants to mark the times that the CMEs pass through two different circles, in the same way they were asked to mark the midpoint time (see Figure 1). This would allow the speeds and accelerations to be calculated with smaller errors.

The SEEDS and CACTus position angles were found to match very well with the SSW position angles; the mean differences were  $-1^\circ$  and  $0^\circ$ , respectively. The SSW errors were between  $1^\circ$  and  $11^\circ$  and therefore larger than the differences between the catalogues. *Robbrecht and Berghams* [2004] found that CACTus position angles matched well with CDAW position angles; as the SSW-CACTus differences were very small it can be inferred that the Solar Stormwatch project was successful in isolating position angles of CMEs.

The matched event list lacked narrow CMEs with widths below  $30^\circ$ , which implies that it has a similar limitation to that found in the expert-produced CDAW catalogue in identifying narrow events [Olmedo *et al.*, 2008]. Furthermore, COR1 and COR2 movies analyzed in SSW were of lower resolution than the raw COR2 and COR1 images, to make it practical to stream the movies over the internet. This reduced resolution probably made it difficult to identify narrow or faint features in the images. Comparison of the SSW, SEEDS, and CACTus widths found that the mean SSW-CACTus width difference was  $8^\circ$  and the mean SSW-SEEDS width difference was  $-19^\circ$ , but these differences could mostly be due to the uncertainty in the SSW widths.

Speeds were calculated by assuming that there was a constant speed between the appearance and midpoint times in the COR1 and COR2 images. However, it was found that CMEs tended to accelerate between the COR1 and COR2 field of view, with an average net speed increase of  $204 \text{ km s}^{-1}$ . Taking account of the uncertainties in the speed calculations, it was shown that all events appeared to accelerate, with only 6 of the 41 CMEs having an uncertainty on the speed increase that could be consistent with a negative acceleration between the COR1 and COR2 fields of view. Therefore, the speeds calculated from the COR1 and COR2 data are more correctly interpreted as average speeds. The COR2 SSW speeds were found to be, on average,  $397 \text{ km/s}$  slower than the SEEDS speeds and  $43 \text{ km/s}$  slower than the CACTus speeds. However, the SSW uncertainties ranged between  $27$  and  $533 \text{ km/s}$  due to the large errors on the appearance times, and therefore any differences in the catalogues could have been due to these errors. *Gopalswamy et al.* [2009] found the mean CME speed to be  $475 \text{ km/s}$ ; however, all but six of the SSW CMEs had speeds below this value, and the mean SSW speed was  $280 \text{ km/s}$ , suggesting that Solar Stormwatch found mostly slow CMEs. Also, *Yashiro et al.* [2008] compared the CDAW and CACTus catalogues and found that CACTus found more CMEs above  $1000 \text{ km/s}$ . The SSW speeds were all below  $1000 \text{ km/s}$ , but this is likely due to the modest number of events identified and that the period of study coincided with generally low solar activity. The HELCATS catalogue of CMEs created manually by an expert will be available for comparison in the near future [EU HELCATS *et al.*, 2015], which will allow direct comparison between these methods.

The CMEs in the SSW event list were found to deflect an average of  $10^\circ$  between the COR1 and COR2 fields of view. Due to projection effects, CME deflections can appear differently depending on the viewpoint of the spacecraft observing them, meaning that there could be biases in the magnitude of the deflections. For this reason, the EUVI data were used to identify limb events, which were analyzed separately as they are simpler to interpret. Four of the CME source regions were off the solar disk, so the nearest location on the disk was used, which could also have caused biases. Plane-of-sky latitude deflection does not give a true representation of the 3-D trajectory of CME deflections; it is possible that the SSW CMEs were deflected in terms of longitude as well as latitude. Due to the viewpoint of the spacecraft, it is possible that the observed deflections might have been deformations of the CME shape. Further work might include looking at case studies of these events to assess whether these deflections did occur. The MAS MHD model [Linker *et al.*, 1999; Riley *et al.*, 2006] was used to simulate the location of the HCS, and the results of the CME deflection analysis depend on the uncertainties inherent in coronal MHD modeling. We tried to mitigate the modeling uncertainties by assessing how robust the result would be to a shift of CME source location. This was done by assuming a  $\pm 10^\circ$  error in the CME source location, computing the HCS location at these limits too, giving a range of HCS locations which were assumed to be a reasonable estimate for the uncertainty in the HCS location.

The conclusion of the deflections analysis was that in general, the SSW CMEs did deflect toward the HCS; only 6 of 34 all-disk events and 4 of 18 limb events did not show deflections toward the HCS. Of the six events

which did not appear to deflect toward the HCS, the uncertainties in the deflections of all but two of these events were large enough that this could have been due to the errors. This result supports the conclusions found by *Shen et al.* [2011], *Gui et al.* [2011], and *Liewer et al.* [2015]. These other studies chose a few specific CMEs to study, whereas here we have considered a larger sample of CMEs, albeit in less detail. This conclusion is potentially useful from a space weather forecasting perspective, as it further demonstrates that CME deflections could cause non-Earth-directed CMEs to become geoeffective as their trajectory changes.

The main results of the paper can be summarized as follows:

1. Using data from the Solar Stormwatch project, we have produced a catalogue of 41 CMEs which tracks events through the COR2, COR1, and EUVI imagers on the STEREO spacecraft.
2. Citizen scientists were found to be effective at isolating the time the CME reaches halfway through the COR1 and COR2 fields of view but less effective at isolating the appearance time.
3. The citizen scientists also effectively isolated the position angles of the events in the COR1 and COR2 fields of view, with only small differences between the SSW, SEEDS, and CACTus CME catalogues on average.
4. The CMEs were found to increase in width between the COR1 and COR2 fields of view, although uncertainties on the width estimates were much larger than those on the position angle estimates.
5. On average, The SSW CMEs were found to accelerate by  $204 \pm 127 \text{ km s}^{-1}$  between the COR1 and COR2 fields of view.
6. CME speeds were found to increase nonlinearly through the EUVI, COR1, and COR2 fields of view.
7. The SSW CMEs were found to deflect by an average of  $10^\circ$  latitude between the COR1 and COR2 fields of view.
8. Of the 36 SSW CMEs for which we had EUVI observations, 28 CMEs were estimated to deflect toward the heliospheric current sheet.

Further work to improve the research presented in this paper could include studying the CME deflections in detail to investigate whether these were actually deflections or deformations of the CME shape. Additionally, the recently released HELCATS catalogue of CMEs could be used to directly compare manual identification of CMEs by an expert with identification of CMEs using citizen scientists. Doing so would give us a better understanding of both the advantages and the limitations of each approach. Improved understanding of the limitations of expert identification of CMEs may be useful for improving space weather forecasting practice, while better understanding of the limitations of the SSW approach will help us develop improved methods for using the citizen scientists' classifications. In particular, this study only used a relatively small percentage of the SSW classifications, and a future study might explore ways to use the classifications more efficiently.

In future citizen science projects, uncertainties in estimates could be reduced by optimizing the user interface to better isolate the CME timing or by using images of higher resolution, which would be possible due to the improvements in technology since the SSW project was started. The citizen scientist method could easily be used to characterize CMEs in data from various different imagers, such as the LASCO instrument on board the Solar and Heliospheric Observatory (SOHO) spacecraft, to create a comprehensive catalogue of CMEs, all classified in the same way. This would be impractical for an expert, due to the time-consuming nature of the task. Extreme ultraviolet images could be obtained from the Atmospheric Imaging Assembly instrument on the Solar Dynamics Observatory (SDO) spacecraft, which have higher resolution than the STEREO EUVI images, but due to the large file sizes, it might be a technical challenge for the computer system of the average citizen scientist to upload many of these images for analysis. However, a better use of citizen scientists might be to focus on more targeted research questions, such as analyzing remote sensing images taken by the upcoming Solar Orbiter mission. The novel observations and limited observation windows that Solar Orbiter will provide might be of more interest to citizen scientists than the long-term cataloguing provided by Solar Stormwatch with STEREO.

## References

- Barnard, L., et al. (2014), The Solar Stormwatch CME catalogue: Results from the first space weather citizen science project, *Space Weather*, 12, 657–674, doi:10.1002/2014SW001119.
- Barnard, L., C. J. Scott, M. Owens, M. Lockwood, K. Tucker-Hood, J. Wilkinson, B. Harder, and E. Baeten (2015a), Solar Stormwatch: Tracking solar eruptions, *Astron. Geophys.*, 56(4), 4.20–4.24, doi:10.1093/astrogeo/atv131.
- Barnard, L., C. J. Scott, M. Owens, M. Lockwood, S. R. Crothers, J. A. Davies, and R. A. Harrison (2015b), Differences between the CME fronts tracked by an expert, an automated algorithm, and the Solar Stormwatch project, *Space Weather*, 13, 709–725, doi:10.1002/2015SW001280.
- Cannon, P., et al. (2013), Extreme space weather: Impacts on engineered systems and infrastructure, *Tech. Rep.*, Royal Acad. of Eng., London.

## Acknowledgments

The Solar Stormwatch classifications and resulting CME catalogue can be found at [https://figshare.com/articles/Solar\\_Stormwatch\\_Track-it-back\\_Catalogue/5224936](https://figshare.com/articles/Solar_Stormwatch_Track-it-back_Catalogue/5224936), and the code used to create and analyze the catalogue is available at <https://github.com/S-hannon/solar-stormwatch-track-it-back>. S.R.J. and L.A.B. thank the Science and Technology Facilities Council for support under grant ST/M000885/1. This research has been made possible by the citizen scientists who contributed to the Solar Stormwatch project. This research has made use of SunPy, an open-source and free community-developed solar data analysis package written in Python [SunPy Community et al., 2015]. All model data can be downloaded from <http://www.predsci.com/mhdweb/home.php>. We thank the STEREO instrument team at the Rutherford Appleton Laboratory for providing the data.

- DeForest, C. E., T. A. Howard, and S. J. Tappin (2011), Observations of detailed structure in the solar wind at 1 AU with STEREO/HI-2, *Astrophys. J.*, **783**, 103–115, doi:10.1088/0004-637X/738/1/103.
- Driesman, A., S. Hynes, and G. Cancro (2007), The STEREO observatory, *Space Sci. Rev.*, **136**, 17–44, doi:10.1007/s11214-007-9286-z.
- Druckmullerova, H., H. Morgan, and S. R. Habbal (2011), Enhancing coronal structures with the Fourier normalizing-radial-graded filter, *Astrophys. J.*, **737**, 88–98, doi:10.1088/0004-637X/737/2/88.
- Gopalswamy, N., S. Yashiro, G. Michalek, G. Stenborg, A. Vourlidas, S. Freeland, and R. Howard (2009), The SOHO/LASCO CME catalog, *Earth Moon Planets*, **104**, 295–313, doi:10.1007/s11038-008-9282-7.
- Gui, B., C. Shen, Y. Wang, P. Ye, J. Liu, S. Wang, and X. Zhao (2011), Quantitative analysis of CME deflections in the corona, *Sol. Phys.*, **271**, 111–139, doi:10.1007/s11207-011-9791-9.
- Hapgood, M. A. (2011), Towards a scientific understanding of the risk from extreme space weather, *Adv. Space Res.*, **47**, 2059–2072, doi:10.1016/j.asr.2010.02.007.
- Harrison, R. G. (2015), *Meteorological Measurements and Instrumentation*, 253 pp., John Wiley, Chichester, U. K.
- EU HELCATS, D. Barnes, J. Byrne, J. Davies, and R. Harrison (2015), HELCATS HCME WP2 V02, doi:10.6084/m9.figshare.1492351.v1.
- Howard, R. A., et al. (2008), Sun Earth Connection Coronal and Heliospheric Investigation (SECCHI), *Space Sci. Rev.*, **136**, 67–115, doi:10.1007/s11214-008-9341-4.
- Kay, C., M. Opher, and R. M. Evans (2015), Global trends of CME deflections based on CME and solar parameters, *Astrophys. J.*, **805**, 168–188, doi:10.1088/0004-637X/805/2/168.
- Liewer, P., O. Panasenco, A. Vourlidas, and R. Colaninno (2015), Observations and analysis of the non-radial propagation of coronal mass ejections near the Sun, *Sol. Phys.*, **290**, 3343–3364, doi:10.1007/s11207-015-0794-9.
- Linker, J. A., Z. Mikic, D. A. Biesecker, R. J. Forsyth, S. E. Gibson, A. J. Lazarus, A. Lecinski, P. Riley, A. Szabo, and B. J. Thompson (1999), Magnetohydrodynamic modeling of the solar corona during Whole Sun Month, *J. Geophys. Res.*, **104**, 9809–9830, doi:10.1029/1998JA900159.
- Morgan, H., J. P. Byrne, and S. R. Habbal (2012), Automatically detecting and tracking coronal mass ejections. I. Separation of dynamic and quiescent components in coronagraph images, *Astrophys. J.*, **752**, 144–158, doi:10.1088/0004-637X/752/2/144.
- Olmedo, O., J. Zhang, H. Wechsler, A. Poland, and K. Borne (2008), Automatic detection and tracking of coronal mass ejections in coronagraph time series, *Sol. Phys.*, **248**, 485–499, doi:10.1007/s11207-007-9104-5.
- Riley, P., J. A. Linker, and Z. Mikic (2006), A comparison between global solar magnetohydrodynamic and potential field source surface model results, *Astrophys. J.*, **653**, 1510–1516, doi:10.1086/508565.
- Riley, P., R. Lionella, J. A. Linker, Z. Mikic, J. Luhmann, and J. Wijaya (2011), Global MHD modeling of the solar corona and inner heliosphere for the whole heliosphere interval, *Sol. Phys.*, **274**, 361–377, doi:10.1007/s11207-010-9698-x.
- Robbrecht, E., and D. Berghams (2004), Automated recognition of coronal mass ejections (CMEs) in near-real-time data, *Astron. Astrophys.*, **425**, 1097–1106, doi:10.1051/0004-6361:20041302.
- Robbrecht, E., D. Berghams, and R. A. M. Van der Linden (2009), Automated LASCO CME catalog for solar cycle 23: Are CMEs scale invariant?, *Astrophys. J.*, **691**, 1222–1234, doi:10.1088/0004-637X/691/2/1222.
- Schrijver, C. J., and G. L. Siscoe (2009), *Prologue. Heliophysics: Plasma Physics of the Local Cosmos*, pp. 1–20, Cambridge Univ. Press, London.
- Shen, C., Y. Wang, B. Gui, P. Ye, and S. Wang (2011), Kinematic evolution of a slow CME in corona viewed by STEREO-B on 8 October 2007, *Sol. Phys.*, **269**, 389–400, doi:10.1007/s11207-011-9715-8.
- Smith, E. J. (2001), The heliospheric current sheet, *J. Geophys. Res.*, **106**, 15,819–15,831, doi:10.1029/2000JA000120.
- SunPy Community et al. (2015), SunPy—Python for solar physics, *Comput. Sci. Discovery*, **8**, 014009, doi:10.1088/1749-4699/8/1/014009.
- Tucker-Hood, K., et al. (2015), Validation of a priori CME arrival predictions made using real-time heliospheric imager observations, *Space Weather*, **13**, 35–48, doi:10.1002/2014SW001106.
- Webb, D. F., and T. A. Howard (2012), Coronal mass ejections: Observations, *Living Rev. Sol. Phys.*, **9**, 9, doi:10.12942/lrsp-2012-3.
- Wilks, D. (2011), *Statistical Methods in the Atmospheric Sciences*, 704 pp. 3rd ed., Academic Press, San Diego, Calif.
- Yashiro, S., G. Michalek, and N. Gopalswamy (2008), A comparison of coronal mass ejections identified by manual and automatic methods, *Ann. Geophys.*, **26**, 3103–3112, doi:10.5194/angeo-26-3103-2008.