Comparing the *Heliospheric Cataloging, Analysis, and Techniques Service* (HELCATS) Manual and Automatic Catalogues of Coronal Mass Ejections Using *Solar Terrestrial Relations Observatory/Heliospheric Imager* (STEREO/HI) Data

L. Rodriguez[1] · D. Barnes[2] · S. Hosteaux[3] · J.A. Davies[2] · S. Willems[1] · V. Pant[4] · R.A. Harrison[2] · D. Berghmans[1] · V. Bothmer[5] · J.P. Eastwood[6] · P.T. Gallagher[7,8] · E.K.J. Kilpua[9] · J. Magdalenic[1,3] · M. Mierla[1,10] · C. Möstl[11] · A.P. Rouillard[12] · D. Odstrčil[13] · S. Poedts[3,14]

Abstract

We present the results of a comparative study between automatic and manually compiled coronal mass ejection (CME) catalogues based on observations from the *Heliospheric Imagers* (HIs) onboard NASA's *Solar Terrestrial Relations Observatory* (STEREO) spacecraft. Using the *Computer Aided CME Tracking software* (CACTus), CMEs are identified in HI data using an automatic feature-detection algorithm, while the *Heliospheric Imagers Catalogue* (HICAT) includes CMEs that are detected by visual inspection of HI images. Both catalogues were compiled as part of the EU FP7 *Heliospheric Cataloguing, Analysis and Techniques Service* (HELCATS) project (www.helcats-fp7.eu). We compare observational parameters of the CMEs from CACTus to those listed in HICAT, such as CME frequency, position angle (PA), and PA-width. We also compare CACTus-derived speeds to speeds derived from applying geometric modelling to the majority of the HICAT CMEs, the results of which are listed in the HELCATS *Heliospheric Imagers Geometric Catalogue* (HIGeoCAT). We find that both CACTus and HICAT catalogues contain a similar number of events when we exclude events narrower than 20°, which are not included in the HICAT catalogue but are found to be identified by CACTus. PA-distributions are strongly peaked around 90° and 270°, with a slightly larger CME frequency northwards of the equatorial plane (particularly for the STEREO-A versions of both catalogues). The CME PA-widths in both HICAT and CACTus catalogues peak at approximately 60°. Manually derived speeds from HIGeoCAT and automatically derived speeds by CACTus correlate well for values lower than 1000 km s^{-1}, in particular when CMEs are propagating close to the plane of the sky.

Keywords Coronal mass ejections · Heliosphere · Space weather

1. Introduction

The advent of wide-angle imaging of the inner heliosphere has revolutionised the study of the solar wind and, in particular, coronal mass ejections (CMEs: e.g. Webb and Howard,

2012) and stream/co-rotating interaction regions (SIRs/CIRs: Belcher and Davis, 1971; Pizzo, 1978; Richardson, 2018). CMEs comprise enormous plasma and magnetic-field structures that are ejected from the Sun and propagate through interplanetary space, while SIRs are characterised by extensive swathes of compressed plasma and magnetic field that form when the fast solar wind (emanating from coronal holes) catches up with the slow solar wind ahead of it. CIRs are SIRs that persist for more than one complete rotation period. CMEs are the main source of intense space-weather effects at the Earth (Gosling, 1993; Koskinen and Huttunen, 2006).

The zodiacal-light photometers onboard the *Helios* spacecraft (Leinert et al., 1975) were used to infer the propagation of solar disturbances into the interplanetary medium in the 1970s and 1980s (e.g. Jackson, 1985; Webb and Jackson, 1990). These observations paved the way for the more recent developments of wide-angle heliospheric imagers, starting with the *Solar Mass Ejection Imager* (SMEI: Eyles et al., 2003) onboard *Coriolis*, followed by the *Heliospheric Imager* (HI: Eyles et al., 2009) onboard each of the pair of *Solar Terrestrial Relations Observatory* (STEREO) spacecraft. A new perspective on the inner heliosphere was then provided by these instruments. Data from STEREO/HI have revealed the nature of solar-wind activity over a region of the heliosphere centred on the ecliptic plane and extending out from around 4° to 90° elongation; this provides a unique resource for investigating CME and SIR/CIR evolution. As STEREO has moved around the Sun, it has provided vantage points away from the Sun–Earth line, which have allowed the HI instruments to clearly observe, in particular, Earth-directed CMEs. At the time of writing, only STEREO-A is operational, since STEREO-B lost communication with the Earth in 2014. The STEREO/HI instruments, in the same way as similar heliospheric imagers on previous (*Coriolis*) and subsequent missions (*Parker Solar Probe* and *Solar Orbiter)*, detect Thomson-scattered light in order to observe CMEs and other solar-wind disturbances, from close to the Sun and up to, and beyond, 1 AU.

The most comprehensive view of CMEs has been obtained from observations by the *Large Angle Spectroscopic Coronagraph* (LASCO: Brueckner et al., 1995) coronagraphs onboard the *Solar and Heliospheric Observatory* (SOHO) spacecraft. These events are listed in the SOHO/LASCO CME catalogue (Gopalswamy et al., 2009), which contains CMEs from 1996 onwards, manually incorporated by a human operator. In an effort to overcome some of the problems related to manual catalogues, Robbrecht and Berghmans (2004) introduced the *Computer Aided CME Tracking software* (CACTus), which analyses the LASCO data and produces a full catalogue of CMEs in an autonomous manner, without human intervention. Both manual and automatic catalogues include relevant CME properties such as position angle (PA), width (or extent), speed, etc. Previous works have dealt with the analysis of these parameters (e.g. Yashiro et al., 2004; Cremades and St. Cyr, 2007, and references therein). Furthermore, several works have been aimed at the comparison of manual and automatic catalogues (e.g. Yashiro, Michalek, and Gopalswamy, 2008; Wang and Colaninno, 2014; Robbrecht, Berghmans, and Van der Linden, 2009; Lamy et al., 2019), with results that have been used in order to further infer properties of CMEs. CACTus has also been applied to the COR2 coronagraph onboard STEREO (Howard et al., 2008), and now to STEREO/HI (Pant et al., 2016, and this work), and these CACTus catalogues are available at www.sidc.be/cactus.

In the framework of the three-year (2014 – 2017) EU-funded *Heliospheric Cataloguing Analysis and Techniques Service* (HELCATS) project, both manual and automatic CME catalogues based on STEREO/HI observations have been created. The HICAT catalogue (Harrison et al., 2018) is based on visual detection of CMEs in images taken by the inner (HI-1) cameras of the HI instruments onboard STEREO. It contains information on the date and

time of first identification of the CME in the HI-1 images, and its PA-limits and -extent. Kinematic properties were then added with the creation of the HIGeoCAT catalogue (Barnes et al., 2019), based on the application of geometrical modelling to the majority of HICAT CMEs. Barnes et al. (2020) subsequently used these catalogues to identify 273 CMEs that were observed simultaneously by the HI instruments on both spacecraft and applied the *Stereoscopic Self-Similar Expansion* (SSSE) technique (Davies et al., 2013) to infer their propagation throughout the inner heliosphere.

CMEs are intrinsically difficult to identify and track in heliospheric-imager data. Furthermore, to do so in an automated way represents an even bigger challenge. Webb and Jackson (1990) created the first catalogue of CMEs in the interplanetary medium, using *Helios-2* data. This catalogue was used to establish their heliospheric characteristics, to compare them with those of CMEs observed near the Sun, and to determine the frequency of occurrence of CMEs from 1976 through 1979. Tappin et al. (2012) created a CME catalogue in heliospheric-imager data using SMEI data and the *Automated Interplanetary Coronal Mass Ejection Detection* (AICMED) technique, however some features in the data led to a high number of false detections. Barnard et al. (2015) reported on a semi-automatic method (*J-tracker*) using STEREO/HI data. Kirnosov, Chang, and Pulkkinen (2016) presented results in automatic tracking of CMEs using COR2 and HI-1 images. All of these methods were applied to a reduced set of events. Under the auspices of HELCATS, Pant et al. (2016) successfully applied an automatic CME detection method to images from the STEREO/HI-1 cameras. This method consists of a modified version of the CACTus algorithm (Robbrecht and Berghmans, 2004). We have now applied this automatic method to the STEREO/HI-1 data available from 2007 until August 2020, creating the first automatic catalogue of CMEs in STEREO/HI data (we call it hereafter the CACTus catalogue for simplicity). In this work, we analyse the CME properties contained within this automatic catalogue and compare them with the same properties contained in the manual CME catalogues HICAT and HIGeoCAT, described above. In this way, we aim at validating the extension of the CACTus technique into the heliospheric-imagers domain. This will be done by comparing an extended set of events, covering more than 13 years of data. The comparison of manual and automatic catalogues is important in order to assess the performance of the event-identification techniques used, which can prove helpful not only for the analysis of past data but also for future applications.

In the next section, the catalogues will be described. In Section 3 we will explain the results found, and in Section 4 conclusions will be drawn.

2. Catalogue Descriptions

The twin-spacecraft of the STEREO mission are in near heliocentric ecliptic orbits, with the STEREO-A(head) spacecraft leading the Earth in its orbit by around 22.5° per year and the STEREO-B(ehind) spacecraft lagging Earth in its orbit by a comparable amount. The HI instruments on STEREO provide wide-angle white-light imaging of the heliosphere out to 1 AU and beyond (Eyles et al., 2009; Harrison et al., 2009). The HI instrument on each STEREO spacecraft comprises two cameras: HI-1 and HI-2. The angular fields of view (FOVs) of HI-1 and HI-2 are 20° and 70°, respectively. With boresights of 14° and 53° elongation in the ecliptic plane for HI-1 and HI-2, respectively, the combined FOV extends from 4° to almost 90° near the Ecliptic in nominal operations. The extensive HI FOV allows us to observe CMEs propagating over vast distances of interplanetary space (Davis et al., 2009; Harrison et al., 2009).

The kinematic properties of manually detected CMEs are listed in the HELCATS HI-GeoCAT catalogue of Barnes et al. (2019), which is an extension of the HICAT catalogue (Harrison et al., 2018). HICAT comprises a list of every CME that is identified in HI-1 images, excluding events that are narrower than 20° in PA. HIGeoCAT was compiled by applying further analysis to a large subset of the HICAT CMEs to determine their kinematic properties based on combined HI-1 and HI-2 data. The same CME kinematic properties as in HIGeoCAT are also included in the automatic CACTus catalogue, albeit calculated in a slightly different manner. That said, for both HIGeoCAT and CACTus speed is determined by tracking the CME through multiple frames taken by HI.

The speeds in the manual HIGeoCAT catalogue are determined using three different geometric-fitting techniques (Davies et al., 2012), which all assume that CME propagation speed and direction are constant and that the CME retains a certain morphology. To generate HIGeoCAT, the CME is manually tracked using time–elongation maps (J-maps: Sheeley et al., 1999, 2008) constructed from HI-1 and HI-2 data along a fixed PA. The selected PA corresponds to the leading part of the eruption, except in the case of faint CMEs where the PA corresponding to the most visible part is used in preference.

The three fitting techniques used to generate HIGeoCAT are based on the fixed-ϕ (FP: Kahler and Webb, 2007; Rouillard et al., 2008; Sheeley et al., 2008), the self-similar expansion geometries (SSE: Davies et al., 2012; Möstl and Davies, 2013), and the harmonic-mean (HM: Lugaz, Vourlidas, and Roussev, 2009; Lugaz et al., 2010) geometries. The first technique considers the CME as a radially moving point source, the third as a radially propagating expanding circle still attached to the Sun, and the second is an intermediate case between the first and third, with the CME assumed to be a radially propagating circle expanding with constant half-width [λ] of 30°. In fact, all three of the geometries can be described by the same SSE equation from Davies et al. (2012):

$$R(t) = v(t - t_0) = \frac{r_0 \sin(\epsilon(t))(1 + \sin(\lambda))}{\sin(\epsilon(t) + \phi) + \sin(\lambda)}, \tag{1}$$

where R is the radial distance of the CME apex from Sun-centre, r_0 the distance between the spacecraft and Sun-centre, ϵ is the elongation angle (the angle from Sun-centre, which corresponds to the tangent line to the assumed circular form of the CME in the case of SSE and HM geometries and to the point-source CME for the FP geometry), ϕ is the spacecraft–Sun–CME angle, and t is time. The half-width angle is 0° for the FP model and 90° for the HM model, by definition, and, as noted above, is set to 30° for the HIGeoCAT implementation of the SSE-based modelling. After manually selecting points coinciding with the leading edge of the CME signature in the J-map constructed from HI-1 and HI-2 data along the selected PA, a fitting algorithm, based on inverting Equation 1, is applied to the resultant time–elongation profile for each of the three assumed geometries to constrain the radial speed and direction relative to the Sun–spacecraft line of the CME apex in the plane defined by the selected PA. As mentioned above, the fitting techniques assume a constant CME radial velocity along a fixed direction. More details on this process can be found in Barnes et al. (2019).

The CME speed in the CACTus catalogue is calculated differently (Pant et al., 2016) from that for HIGeoCAT. Instead of using a specific PA, CACTus considers the CME as a cluster of points, forming a map. Each cluster is associated with a velocity (see Figure 4 in Pant et al., 2016). For each cluster of points in a CME map, and thus for each CME, we identify the maximum velocity at each PA. Therefore, for each CME, CACTus provides upper and lower bounds of the velocity and a median velocity value. By using the expression from

Pant et al. ([2016](#)),

$$R = r_0 \tan\epsilon, \tag{2}$$

we assume that the CME is directed in the plane of the sky (POS), i.e. at 90° from the Sun–spacecraft line. This makes it a special case of the FP version of Equation [1](#).

To compare the geometrical and kinematic properties of CMEs by CACTus with those in HICAT and HIGeoCAT, respectively, it is necessary to identify CMEs common to both catalogues; in the next section we will explain how this was carried out.

3. Results

3.1. CME Frequency

CMEs in both the manual HICAT and automatic CACTus catalogues between 15 April 2007 and 31 August 2020 are considered (see Figure [1](#)). The manual catalogue contains a comparable number of events for both STEREO spacecraft (965 for A and 936 for B), during the period prior to superior conjunction, from April 2007 to September 2014, when both spacecraft were still operational before the loss of STEREO-B. Note that these numbers represent all CMEs in HICAT, including good, fair, and poor events. The automatic catalogue contains approximately 20% more events from STEREO-B (1826) when compared to STEREO-A (1525) during the same period. The greater number of events identified by CACTus in HI-1B images relates to the fact that HI-1B undergoes sporadic movements due to an inadequacy in the mounting of the camera focal-plane assembly (Tappin, [2017](#)). This compromises the background subtraction for some intervals, causing residual background (mainly F-corona) to be identified as CMEs by CACTus, resulting in the greater number of detections seen in Figure [1](#)d. Manual cataloguing is less prone to such misidentifications, but this effect may account for the fact that marginally fewer CMEs are identified by eye in HI-1B images, because the residual background could mask faint and relatively small CMEs (Harrison et al., [2018](#)).

For the entire period under consideration, from 15 April 2007 to 31 August 2020, the total number of events detected in STEREO-A by the automatic method is much higher than in the manual case (2195 vs. 1319, respectively). Note that HICAT was compiled only for events with a PA-extent of at least 20°, and so this difference in the number of events is resolved if we set the automatic algorithm to detect only events with a PA-width larger than 20°. By doing so, the total number of events in the automatic catalogue reduces to 1442 and becomes comparable with the number of events in the manual catalogue during the interval under study. This means that a significant fraction (around one third) of the CMEs identified by CACTus have a narrow angular width (<20°). As noted above, such events are not included in the manual catalogue. The number of CACTus CMEs detected by STEREO-B with a PA-width of at least 20° is 1524 (compared to the total of 1826), which is still much greater than the 936 from HICAT; as discussed above, this discrepancy is due to the erroneous detections made by CACTus when applied to HI-1 data from STEREO-B. For the rest of the article we will only make comparisons between CMEs wider than 20°, and for the time period cited above (15 April 2007 to 31 August 2020).

In order to find events common to both catalogues, we have searched the HICAT and CACTus catalogues for CMEs that first appear in the HI-1 images within a two-hour time window of each other. We consider all HICAT CMEs, including poor events, and exclude CACTus CMEs with a PA-width lower than 20°. If there is more than one CME within that two-hour window, the closest one in time is taken. In this way, we identified 746 events com-
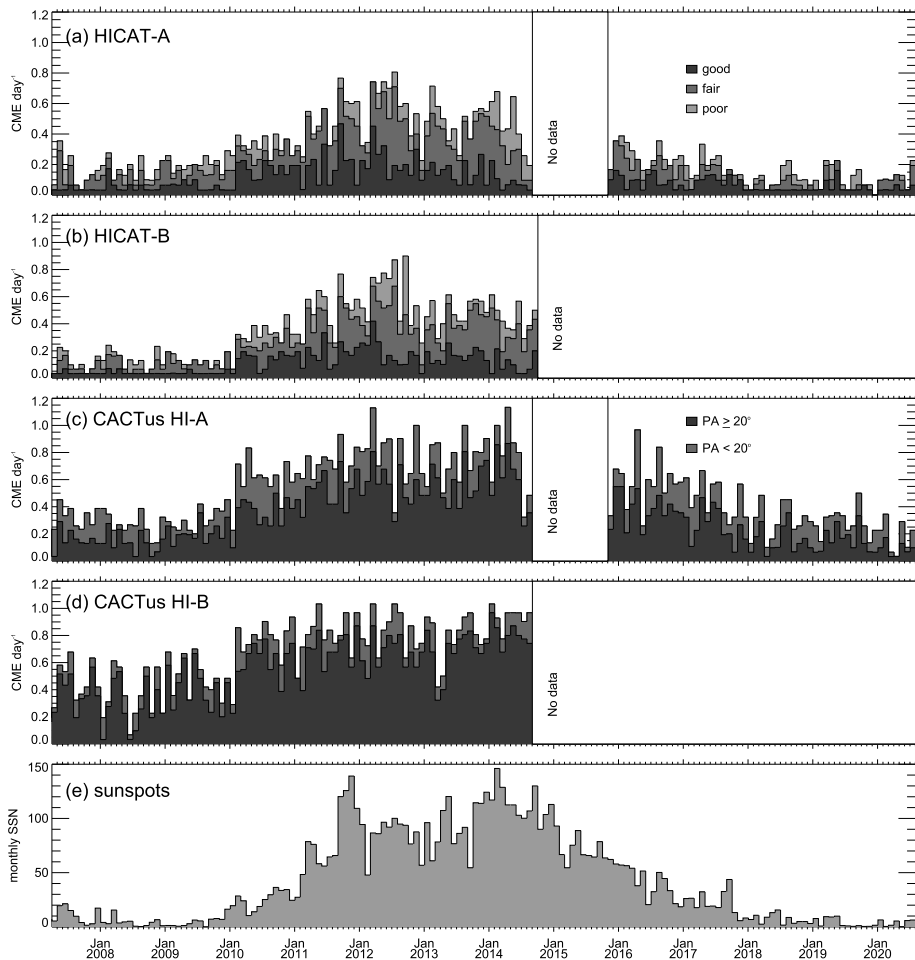
Figure 1 Daily averaged CME count, binned by calendar month from HICAT for STEREO-A (**a**) and -B (**b**) and from CACTus for STEREO-A (**c**) and -B (**d**). The monthly sunspot number is shown in Panel **e** (source: SILSO, Royal Observatory of Belgium). The shaded histograms in Panels **a** and **b** represent the fraction of *good* (*dark grey*), *fair* (*medium grey*), and *poor* (*light grey* events). The *lighter and darker distributions* in Panels **c** and **d** represent CMEs with PA-width $< 20°$ and $\geq 20°$, respectively. Data gaps represent periods where the STEREO spacecraft are out of contact with Earth.

mon to both HI-1A catalogues and 474 common to both HI-1B catalogues. Figure 2 shows an example of a common CME, observed on 25 February 2010 by HI-1 on STEREO-A. The left panel shows the manually detected CME and the right one the automatically identified CME (the pre-processing of the images is different in both cases). Both methods give similar results for the PA-extent of the CME, the northern and southern PAs are 145° and 80°, respectively, in the manual catalogue and 136° and 84° in the automatic one. The angles in the manual catalogue are quoted with a resolution of 5°; for the automated catalogue (as described by Pant et al., 2016) the resolution in angle was kept at 2°.
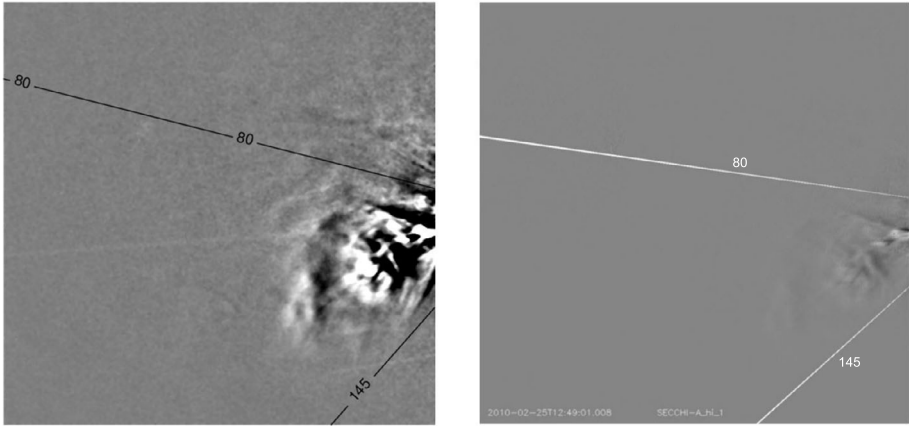
Figure 2 Example of a CME observed in HI-1 data from STEREO-A on 25 February 2010, in the HICAT manual catalogue (*left*) and in the automatic CACTus catalogue (*right*). Both panels extend from the Sun up to 20° of elongation. The *lines* mark the most northern and southern PAs.
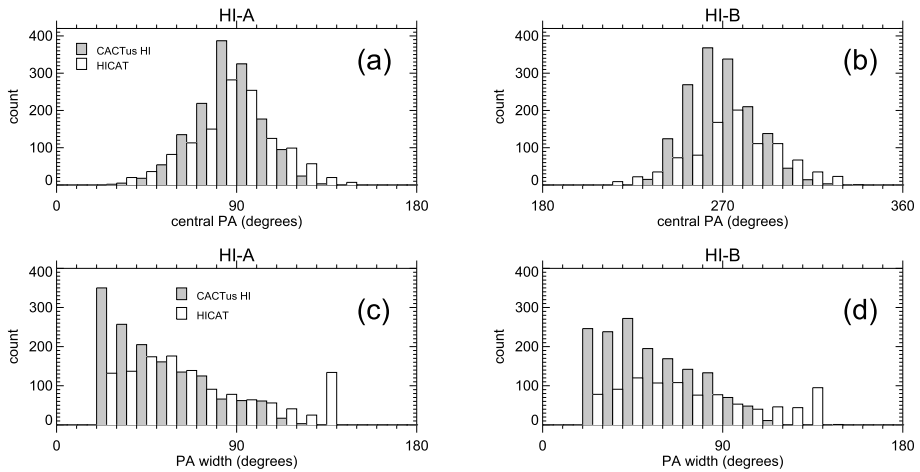
Figure 3 Distributions of CME central PA for CMEs detected by STEREO-A (**a**) and -B (**b**). Distributions of CME PA-width for CMEs detected by STEREO-A (**c**) and -B (**d**). In each panel, *the distributions for HICAT and CACTus are represented by white and grey*, respectively.

3.2. CME Observational Characteristics

Figures 3a and b show the distributions of CME central PA, for both catalogues, from each spacecraft. The central PA is taken as the mid-point between the northern and southern PA. Both the STEREO-A distributions are strongly peaked around 90° and both the STEREO-B distributions peak around 270°. These values correspond to equatorial latitudes, which is consistent with CME behaviour established in a number of existing catalogues (e.g. Howard et al., 1985; Burkepile and St. Cyr, 1993; Yashiro et al., 2004; Gopalswamy et al., 2009) and, indeed, Harrison et al. (2018) in their discussion of HICAT. For those events that extend beyond the northern and/or southern extent of the HI-1 FOV, the northern or southern PA

and the PA-width will be artificially truncated in both the manual and automatic catalogues. Similarly, the central PA of those CMEs, calculated as the mid-point between the upper and lower PA-limits, is likely to be biased towards PAs closer to the Equator than in reality. This effect is particularly significant if only the northern or southern edge of the CME extends beyond the HI-1 FOV. Note that both catalogues exhibit a slightly larger CME frequency northwards of the equatorial plane (except HICAT for STEREO-B).

Harrison et al. (2018) performed an analysis of the manually detected HI-1 CMEs and compared them with CMEs observed in COR-2 coronagraph images, based on the COR-2 catalogue described by Vourlidas et al. (2017). They found that coronal CMEs have a much larger range of central PAs than heliospheric CMEs and that CMEs become more constrained to equatorial PAs (equivalent to lower latitudes) as they propagate outwards from the Sun. The authors suggest that this can be explained by the tendency for CMEs to be driven towards the Equator when interacting with the large-scale solar magnetic field (Plunkett et al., 2001; Kilpua et al., 2009; Byrne et al., 2010; Harrison et al., 2012). The presence of an overlying helmet streamer guides the eruption to lower latitudes, with a stronger helmet streamer's magnetic field resulting in more equatorward deflection (See Zuccarello et al., 2011).

Figures 3c and d show the PA-widths of CMEs in both catalogues. Both the manually identified HICAT STEREO-A and -B CME width distributions peak at approximately 60°, as shown by Harrison et al. (2018). However, for STEREO-A, the CACTus distribution peaks in the narrowest bin (20–30°) and decreases steadily for CMEs of increasing width. These distributions suggest that a number of narrower CMEs may be missed during manual detection, because it is in the narrower bins that the distributions diverge. Similar behaviour was noted when CACTus for LASCO was compared with the CDAW SOHO CME catalogue (Robbrecht, Berghmans, and Van der Linden, 2009). A scale-free nature of CMEs was observed in the PA-widths estimated by CACTus, but this behaviour was not present in the manual catalogue. The STEREO-B CACTus distribution intimates a similar scale-free behaviour although peaking at around 50°, but many of these events are erroneous detections, so it is difficult to draw meaningful conclusions. A second peak is visible at the upper end of the distributions for the HICAT catalogues, which is not observed by CACTus, presumably because CACTus CMEs tend to be narrower. These secondary HICAT peaks, as mentioned above, are due to the artificial limit imposed on those CMEs that are wider than 140°, due to the restricted PA-extent of the HI-1 FOV. There is a lack of wide events (>125°) in the automatic catalogue and an excess of small events (<50°). Pant et al. (2016) found a similar discrepancy while analysing a reduced set of events, including 15 common CMEs. A possible explanation for this is that the flanks of a CME are often fainter than its leading edge, and thus they may not be sufficiently bright to trigger a detection by CACTus. Another possible reason could be that CME shocks may be present in some of the events, and they tend to be wider but fainter than the CME itself, again not triggering the CACTus threshold for detection while potentially being visible to a manual observer. Finally, a known CACTus issue is that it tends to identify a large faint event as several narrower events. As noted above, this discrepancy in angular width is also present when comparing manual with automatic CME detections in coronagraph data (Robbrecht, Berghmans, and Van der Linden, 2009). Lowering the intensity thresholds in CACTus would help to resolve this problem, but it would also increase the number of false detections.

Figure 4 shows a direct comparison of the CME PA-widths from the distributions in Figure 3, only for CMEs that were matched between the two catalogues. We see a good correlation in both panels of Figure 4, particularly for CMEs detected in HI-A. The lack of wide events (>125°) in the automatic catalogue, and an excess of small events (<50°),
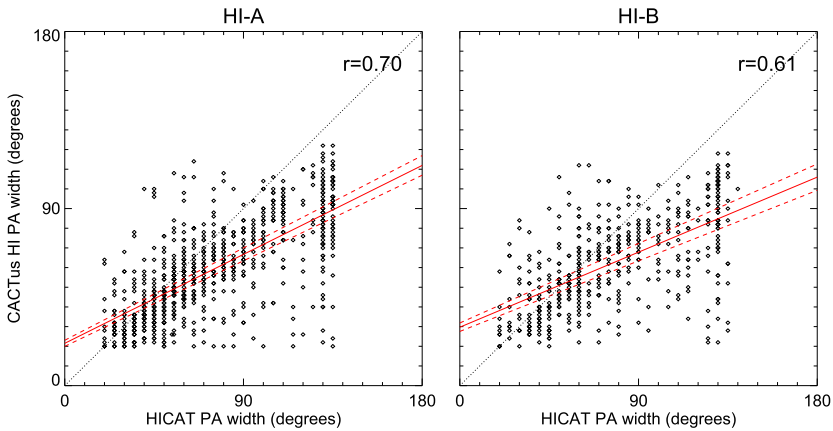
Figure 4 Scatter plots comparing PA-widths of CMEs identified in HICAT and CACTus from STEREO-A (*left*) and STEREO-B (*right*). The regression line between the two data sets is over-plotted in both cases as a *red line* with error bounds indicated by *red-dashed* lines, with the correlation coefficient printed in the *top-right corner*, and the *black-dotted* line is the identity line.

described above, can also be seen in the scatter plot of CME PA-width shown in Figure 4, where the vast majority (82%) of common CME widths are below the diagonal. On the contrary, there are a few events where CACTus is detecting wide events (around 100°) that are seen as narrow events ($\approx 50°$) in the manual HICAT catalogue.

3.3. CME Kinematics

In Figures 5, 6, and 7 we compare the speeds of the CMEs in CACTus and HIGeoCAT. The latter contains a large subset of the HICAT CMEs, in particular those that were classified as *fair* or *good* and excluding the *poor* events, or ones that could not be clearly tracked throughout the combined HI-1 and HI-2 FOVs in order to record their observed elongation as a function of time (Barnes et al., 2019). The comparison of speed between CACTus and HIGeoCAT has been performed for matching events (using the same criteria for the matching of events between CACTus and HICAT). The speed has been calculated using each of the FP, SSE, and HM fitting techniques for the manually generated catalogue. This refers to a radial speed, whereas CACTus always gives a POS speed. The results from the three fitting methods do not vary significantly, although the difference depends on the direction of the CME relative to the Sun–spacecraft line.

In each case, the comparison between manual and automatically derived speeds using STEREO-A data is good for low speeds (less than 1000 km s^{-1}). For larger speed values, the correlation is poor. This is a consequence of the fact that higher-speed CMEs result in fewer images in which the CME is present, and therefore fewer data points with which to fit Equations 1 and 2. This increases the errors in the measurements of both the manual and the automatic methods, however, the manual method uses both HI-1 and HI-2 data, so it is less susceptible to this error. Nevertheless, HIGeoCAT requires a longer track in order to be accurate, because it uses the curvature to provide an estimate of speed and direction. Therefore, a short track will give erroneous results due to degeneracy. These effects are apparent in Figures 5, 6, and 7, where, in all the scatter plots, the values based on shorter tracks (blue) tend to lie further from the identity line than those based on the longest tracks (red). For
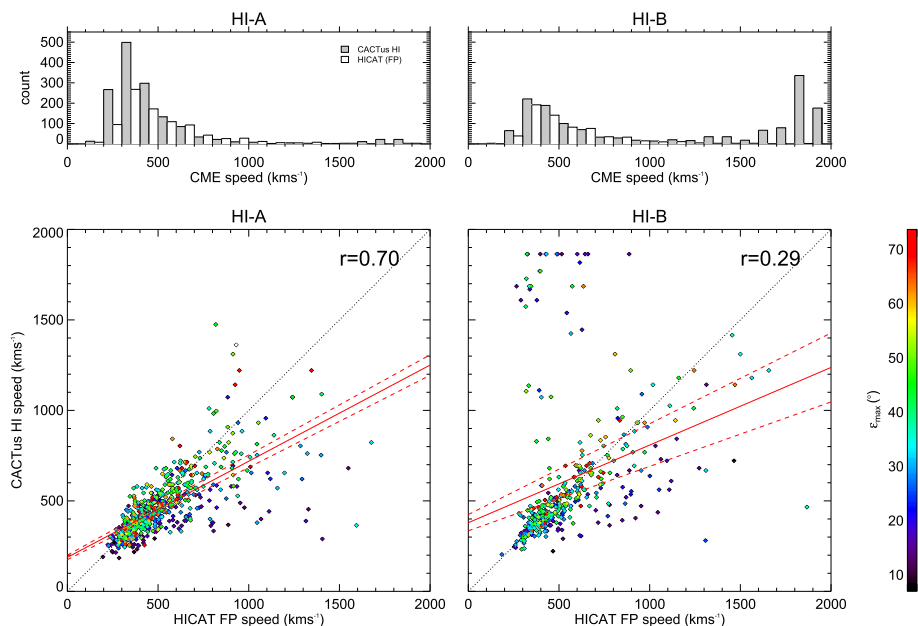
Figure 5 *The top panels* show the distributions of CME speeds from CACTus (*grey*) and speeds derived using FP fitting from HIGeoCAT (*white*); this is done for the two catalogues separately. The *left- and right-hand panels* correspond to STEREO-A and -B, respectively. The *bottom panels* show scatter plots comparing the CACTus and HIGeoCAT speeds for the matched events, with *regression lines over-plotted in red*. *The colour of the symbols represents the maximum elongation* to which the HIGeoCAT CME was tracked, as shown on the colour bar.

STEREO-B, a large number of events detected using CACTus are found to have speeds well in excess of 1000 km s^{-1}. This is a result of trying to fit speeds to erroneously identified CMEs, which causes the distributions to have a second peak at around 1800 km s^{-1}.

Both catalogues show the same general characteristics in their speed distributions for STEREO-A, with CME speeds starting at 200 km s^{-1} and peaking at 350–400 km s^{-1}. In general, CACTus CME speeds tend to be higher than HIGeoCAT speeds between 200 and 800 km s^{-1}, while the opposite is true for higher velocities. The derived speed is dependent on the propagation direction, which is estimated in the manual catalogue but assumed in the POS in the CACTus catalogue. Thus, for ϕ-values close to 90° the estimated velocities should be in better agreement with each other (assuming the derived ϕ is accurate), but if the CME is directed away from the limb, meaning that ϕ is not close to 90°, the POS assumption will lead to an underestimation of the true radial speed, as shown by Barnes et al. (2019). This is also shown in a limited sample analysed by Pant et al. (2016). Similarly, if a shock is detected as the leading part of the CME by the manual observer but not by the automated detection method, the manual velocity would likely be higher than the CACTus velocity. Overall, the FP-fitting technique matches the values from CACTus the best. This is likely because the geometry of the FP-method is equivalent to the geometry assumed for CACTus and, for a given CME, the speed retrieved from the FP-method is invariably lower than those from SSE or HM. The correlation between STEREO-A CME speeds between HIGeoCAT and CACTus is best (0.70) for this case (FP) and worst (0.60) when using the HM fitting. For STEREO-B CMEs, the correlation between speeds determined by HIGeoCAT and CACTus is poor for all of the fitting methods, due to the problem with HI-B CME detections in CACTus.
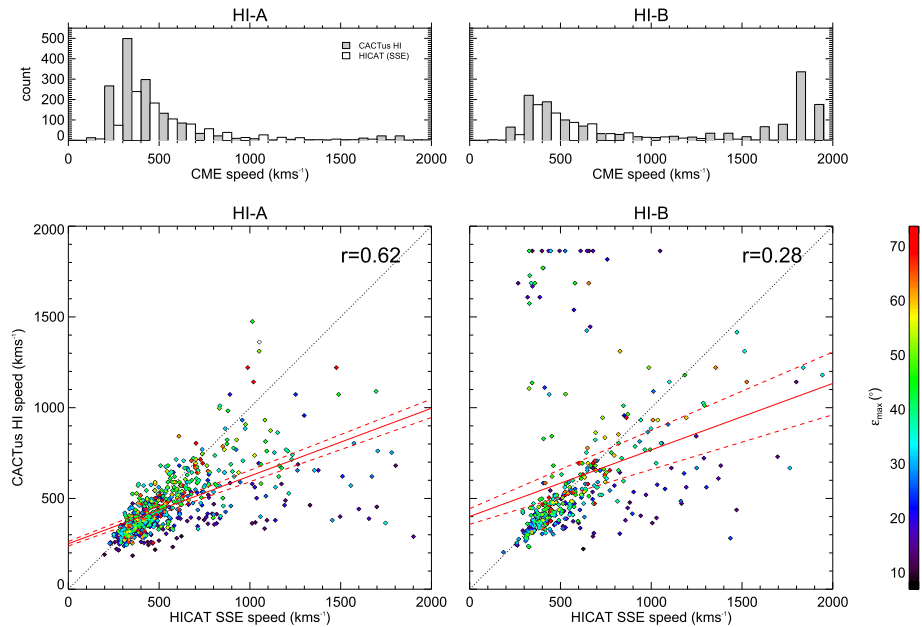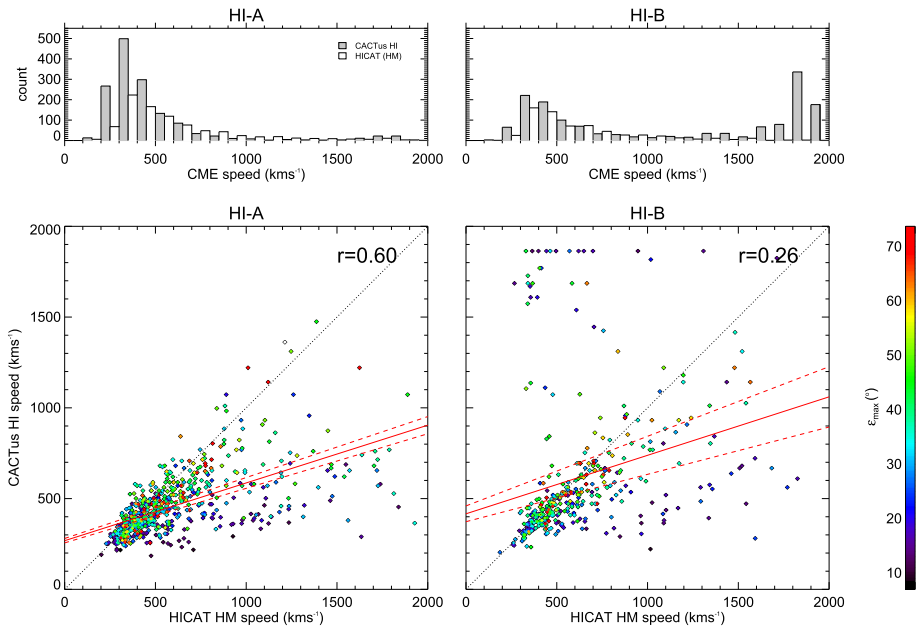
Figure 6 *The top panels* show the distributions of CME speeds from CACTus (*grey*) and speeds derived using SSE fitting with a half-width of 30° from HIGeoCAT (*white*); this is done for the two catalogues separately. The *left- and right-hand panels* correspond to STEREO-A and -B, respectively. *The bottom panels* show scatter plots comparing the CACTus and HIGeoCAT speeds for the matched events, *with regression lines over-plotted in red. The colour of the symbols represents the maximum elongation* to which the HIGeoCAT CME was tracked, as shown on the colour bar.

4. Summary and Conclusions

A comparative study was performed between STEREO/HI CME catalogues, namely the HI-CAT and HIGeoCAT manual catalogues and the CACTus automatic catalogue. It is important to compare manual and automatic catalogues, in order to assess the quality of the techniques used. Constant validation against data is the most transparent way of validating automatic-detection methods.

The number of recorded events by CACTus is much higher than those recorded in the manual catalogue over the same period. A reason for this is that CACTus often detects a single CME as multiple separate events. Moreover, the manual catalogue employs a 20° minimum-width threshold. Analysis was performed on common CMEs, which are defined as CMEs that appear in both catalogues with less than a two-hour time difference. This results in a list of 746 common CMEs in HI-1A and 474 in HI-1B. There is, in general, a good agreement between both catalogues in terms of the derived CME central PAs and PA-widths. The speeds of CMEs detected using STEREO-A agree well between the manual and automatic catalogues (for speeds lower than $1000 \, \text{km} \, \text{s}^{-1}$); however, the speeds of CMEs detected using STEREO-B data show poor agreement, resulting from the problems with HI-B data that particularly affect CACTus CME detections.

The central-PA distributions in both catalogues are centred around the Equator. For some cases, CMEs will have a central PA listed closer to the Equator than its true value, due to one of its edges extending beyond the FOV of HI-1. The difference in CME central PA between

Figure 7 *The top panels* show the distributions of CME speeds from CACTus (*grey*) and speeds derived using HM fitting from HIGeoCAT (*white*); this is done for the two catalogues separately. The *left- and right-hand panels* correspond to STEREO-A and -B, respectively. The *bottom panels* show scatter plots comparing the CACTus and HIGeoCAT speeds for the matched events, with *regression lines over-plotted in red*. The colour of the symbols represents the maximum elongation to which the HIGeoCAT CME was tracked, as shown on the colour bar.

the two catalogues is small, showing that they are in fairly good agreement. CMEs are listed with PA-widths ranging from 20° to 140°, with PA-widths in both catalogues peaking at around 60°. However, CMEs are more likely to be listed with a smaller apparent width in the CACTus catalogue, due to the CME edges possibly being too faint to be detected by CACTus, or the width of a preceding (wider) shock wave being taken as the CME width by the manual observer. Moreover, structuring of the brightness inside the CME can cause CACTus to list the CME as multiple separate, narrower, events. CME speeds in both catalogues start at around $200 \, \text{km} \, \text{s}^{-1}$, with most events having a speed of approximately $400 \, \text{km} \, \text{s}^{-1}$. Since CACTus assumes the POS approximation for calculating the speeds, the catalogues are more likely to be in agreement for near-limb events (where the direction of the CME relative to the Sun–spacecraft line is close to 90°). The speed of slow CMEs agrees better between the catalogues than faster ones; this is a consequence of having fewer frames to calculate CME speeds in fast events, leading to larger errors in both catalogues (but mainly in the automatic catalogue, since it uses only HI-1 data, while the manual one also includes HI-2 data in its speed estimation). For CMEs away from the limb, CACTus underestimates the 3D speed found using the fitting techniques in the manual catalogue.

The EU FP7 HELCATS project has created a wealth of relevant catalogues of CME (and CIR)-related science, not only using observations but also physics-based heliospheric models, for example several catalogues were created using the 3D MHD code ENLIL (Odstrcil, Riley, and Zhao, 2004); they can be accessed from the HELCATS website www.helcats-fp7.eu. All of these catalogues remain available for future use. For example, the EU H2020

project *EUropean Heliospheric FORecasting Information Asset* (EUHFORIA) 2.0, among others, is making use of the information collected by HELCATS in order to improve the 3D MHD model EUHFORIA (Pomoell and Poedts, 2018). Furthermore, most of these catalogues are still being constantly updated (CACTus in near-real time), even after the end of the project. These efforts are important in terms of improving and validating operational forecasting services, which tend to become more automated with time. In that sense, they are very relevant for the upcoming ESA *Vigil* mission, which will enable improved forecasts of space-weather phenomena at Earth, through real-time monitoring of the Sun and solar wind from the L_5 vantage point. This marks the importance of projects such as HELCATS.

Declarations

Disclosure of Potential Conflicts of Interest The authors declare that they have no conflicts of interest.

References

Barnard, L., Scott, C.J., Owens, M., Lockwood, M., Crothers, S.R., Davies, J.A., Harrison, R.A.: 2015, Differences between the CME fronts tracked by an expert, an automated algorithm, and the Solar Stormwatch project. *Space Weather* **13**, 709. DOI. ADS.

Barnes, D., Davies, J.A., Harrison, R.A., Byrne, J.P., Perry, C.H., Bothmer, V., Eastwood, J.P., Gallagher, P.T., Kilpua, E.K.J., Möstl, C., Rodriguez, L., Rouillard, A.P., Odstrčil, D.: 2019, CMEs in the heliosphere: II. A statistical analysis of the kinematic properties derived from single-spacecraft geometrical modelling techniques applied to CMEs detected in the heliosphere from 2007 to 2017 by STEREO/HI-1. *Solar Phys.* **294**, 57. DOI.

Barnes, D., Davies, J.A., Harrison, R.A., Byrne, J.P., Perry, C.H., Bothmer, V., Eastwood, J.P., Gallagher, P.T., Kilpua, E.K.J., Möstl, C., Rodriguez, L., Rouillard, A.P., Odstrčil, D.: 2020, CMEs in the heliosphere: III. A statistical analysis of the kinematic properties derived from stereoscopic geometrical modelling techniques applied to CMEs detected in the heliosphere from 2008 to 2014 by STEREO/HI-1. *Solar Phys.* **295**, 150. DOI. ADS.

Belcher, J.W., Davis, L. Jr.: 1971, Large-amplitude Alfvén waves in the interplanetary medium, 2. *J. Geophys. Res. (1896-1977)* **76**, 3534. DOI.

Brueckner, G.E., Howard, R.A., Koomen, M.J., Korendyke, C.M., Michels, D.J., Moses, J.D., Socker, D.G., Dere, K.P., Lamy, P.L., Llebaria, A., Bout, M.V., Schwenn, R., Simnett, G.M., Bedford, D.K., Eyles, C.J.: 1995, The Large Angle Spectroscopic Coronagraph (LASCO). *Solar Phys.* **162**, 357. DOI. ADS.

Burkepile, J.T., St. Cyr, O.C.: 1993, *A Revised and Expanded Catalogue of Mass Ejections Observed by the Solar Maximum Mission Coronagraph, TN-369+STR*, NCAR/UCAR, Boulder. DOI. ADS.

Byrne, J.P., Maloney, S.A., McAteer, R.T.J., Refojo, J.M., Gallagher, P.T.: 2010, Propagation of an Earth-directed coronal mass ejection in three dimensions. *Nat. Commun.* **1**, 74. DOI. ADS.

Cremades, H., St. Cyr, O.C.: 2007, Coronal mass ejections: solar cycle aspects. *Adv. Space Res.* **40**, 1042. DOI. ADS.

Davies, J.A., Harrison, R.A., Perry, C.H., Möstl, C., Lugaz, N., Rollett, T., Davis, C.J., Crothers, S.R., Temmer, M., Eyles, C.J., Savani, N.P.: 2012, A self-similar expansion model for use in solar wind transient propagation studies. *Astrophys. J.* **750**, 23. DOI. ADS.

Davies, J.A., Perry, C.H., Trines, R.M.G.M., Harrison, R.A., Lugaz, N., Möstl, C., Liu, Y.D., Steed, K.: 2013, Establishing a stereoscopic technique for determining the kinematic properties of solar wind transients based on a generalized self-similarly expanding circular geometry. *Astrophys. J.* **777**, 167. DOI. ADS.

Davis, C.J., Davies, J.A., Lockwood, M., Rouillard, A.P., Eyles, C.J., Harrison, R.A.: 2009, Stereoscopic imaging of an Earth-impacting solar coronal mass ejection: a major milestone for the STEREO mission. *Geophys. Res. Lett.* **36**, L08102. DOI. ADS.

Eyles, C.J., Simnett, G.M., Cooke, M.P., Jackson, B.V., Buffington, A., Hick, P.P., Waltham, N.R., King, J.M., Anderson, P.A., Holladay, P.E.: 2003, The solar mass ejection imager (SMEI). *Solar Phys.* **217**, 319. DOI. ADS.

Eyles, C.J., Harrison, R.A., Davis, C.J., Waltham, N.R., Shaughnessy, B.M., Mapson-Menard, H.C.A., Bewsher, D., Crothers, S.R., Davies, J.A., Simnett, G.M., Howard, R.A., Moses, J.D., Newmark, J.S., Socker, D.G., Halain, J.-P., Defise, J.-M., Mazy, E., Rochus, P.: 2009, The heliospheric imagers onboard the STEREO mission. *Solar Phys.* **254**, 387. DOI. ADS.

Gopalswamy, N., Yashiro, S., Michalek, G., Stenborg, G., Vourlidas, A., Freeland, S., Howard, R.: 2009, The SOHO/LASCO CME catalog. *Earth Moon Planets* **104**, 295. DOI. ADS.

Gosling, J.T.: 1993, The solar flare myth. *J. Geophys. Res.* **98**, 18937. DOI. ADS.

Harrison, R.A., Davies, J.A., Rouillard, A.P., Davis, C.J., Eyles, C.J., Bewsher, D., Crothers, S.R., Howard, R.A., Sheeley, N.R., Vourlidas, A., Webb, D.F., Brown, D.S., Dorrian, G.D.: 2009, Two years of the STEREO heliospheric imagers. Invited review. *Solar Phys.* **256**, 219. DOI. ADS.

Harrison, R.A., Davies, J.A., Möstl, C., Liu, Y., Temmer, M., Bisi, M.M., Eastwood, J.P., de Koning, C.A., Nitta, N., Rollett, T., Farrugia, C.J., Forsyth, R.J., Jackson, B.V., Jensen, E.A., Kilpua, E.K.J., Odstrcil, D., Webb, D.F.: 2012, An analysis of the origin and propagation of the multiple coronal mass ejections of 2010 August 1. *Astrophys. J.* **750**, 45. DOI. ADS.

Harrison, R.A., Davies, J.A., Barnes, D., Byrne, J.P., Perry, C.H., Bothmer, V., Eastwood, J.P., Gallagher, P.T., Kilpua, E.K.J., Möstl, C., Rodriguez, L., Rouillard, A.P., Odstrčil, D.: 2018, CMEs in the heliosphere: I. A statistical analysis of the observational properties of CMEs detected in the heliosphere from 2007 to 2017 by STEREO/HI-1. *Solar Phys.* **293**, 77. DOI. ADS.

Howard, R.A., Sheeley, N.R. Jr., Michels, D.J., Koomen, M.J.: 1985, Coronal mass ejections - 1979-1981. *J. Geophys. Res.* **90**, 8173. DOI. ADS.

Howard, R.A., Moses, J.D., Vourlidas, A., Newmark, J.S., Socker, D.G., Plunkett, S.P., Korendyke, C.M., Cook, J.W., Hurley, A., Davila, J.M., Thompson, W.T., St Cyr, O.C., Mentzell, E., Mehalick, K., Lemen, J.R., Wuelser, J.P., Duncan, D.W., Tarbell, T.D., Wolfson, C.J., Moore, A., Harrison, R.A., Waltham, N.R., Lang, J., Davis, C.J., Eyles, C.J., Mapson-Menard, H., Simnett, G.M., Halain, J.P., Defise, J.M., Mazy, E., Rochus, P., Mercier, R., Ravet, M.F., Delmotte, F., Auchere, F., Delaboudinière, J.P., Bothmer, V., Deutsch, W., Wang, D., Rich, N., Cooper, S., Stephens, V., Maahs, G., Baugh, R., McMullin, D., Carter, T.: 2008, Sun Earth connection coronal and heliospheric investigation (SECCHI). *Space Sci. Rev.* **136**, 67. DOI. ADS.

Jackson, B.V.: 1985, Imaging of coronal mass ejections by the HELIOS spacecraft. *Solar Phys.* **100**, 563. DOI. ADS.

Kahler, S.W., Webb, D.F.: 2007, V arc interplanetary coronal mass ejections observed with the Solar Mass Ejection Imager. *J. Geophys. Res.* **112**, A09103. DOI. ADS.

Kilpua, E.K.J., Pomoell, J., Vourlidas, A., Vainio, R., Luhmann, J., Li, Y., Schroeder, P., Galvin, A.B., Simunac, K.: 2009, STEREO observations of interplanetary coronal mass ejections and prominence deflection during solar minimum period. *Ann. Geophys.* **27**, 4491. DOI. ADS.

Kirnosov, V., Chang, L.-C., Pulkkinen, A.: 2016, Combining STEREO SECCHI COR2 and HI1 images for automatic CME front edge tracking. *J. Space Weather Space Clim.* **6**, A41. DOI. ADS.

Koskinen, H.E.J., Huttunen, K.E.J.: 2006, Geoeffectivity of coronal mass ejections. *Space Sci. Rev.* **124**, 169. DOI. ADS.

Lamy, P.L., Floyd, O., Boclet, B., Wojak, J., Gilardy, H., Barlyaeva, T.: 2019, Coronal mass ejections over solar cycles 23 and 24. *Space Sci. Rev.* **215**, 39. DOI. ADS.

Leinert, C., Link, H., Pitz, E., Salm, N., Knueppelberg, D.: 1975, The Helios zodiacal light experiment (E9). *Raumfahrtforschung* **19**, 264. ADS.

Lugaz, N., Vourlidas, A., Roussev, I.I.: 2009, Deriving the radial distances of wide coronal mass ejections from elongation measurements in the heliosphere - application to CME-CME interaction. *Ann. Geophys.* **27**, 3479. DOI. ADS.

Lugaz, N., Hernandez-Charpak, J.N., Roussev, I.I., Davis, C.J., Vourlidas, A., Davies, J.A.: 2010, Determining the azimuthal properties of coronal mass ejections from multi-spacecraft remote-sensing observations with stereo secchi. *Astrophys. J.* **715**, 493.

Möstl, C., Davies, J.A.: 2013, Speeds and arrival times of solar transients approximated by self-similar expanding circular fronts. *Solar Phys.* **285**, 411. DOI. ADS.

Odstrcil, D., Riley, P., Zhao, X.P.: 2004, Numerical simulation of the 12 May 1997 interplanetary CME event. *J. Geophys. Res.* **109**, A02116. DOI. ADS.

Pant, V., Willems, S., Rodriguez, L., Mierla, M., Banerjee, D., Davies, J.A.: 2016, Automated detection of coronal mass ejections in stereo heliospheric imager data. *Astrophys. J.* **833**, 80.

Pizzo, V.: 1978, A three-dimensional model of corotating streams in the solar wind, 1. Theoretical foundations. *J. Geophys. Res.* **83**, 5563. DOI.

Plunkett, S.P., Thompson, B.J., St. Cyr, O.C., Howard, R.A.: 2001, Solar source regions of coronal mass ejections and their geomagnetic effects. *J. Atmos. Solar-Terr. Phys.* **63**, 389. Interplanetary medium and geophysical phenomena during geomagnetic storms. DOI.

Pomoell, J., Poedts, S.: 2018, EUHFORIA: European heliospheric forecasting information asset. *J. Space Weather Space Clim.* **8**, A35. DOI. ADS.

Richardson, I.G.: 2018, Solar wind stream interaction regions throughout the heliosphere. *Liv. Rev. Solar Phys.* **15**, 1. DOI. ADS.

Robbrecht, E., Berghmans, D.: 2004, Automated recognition of coronal mass ejections (CMEs) in near-real-time data. *Astron. Astrophys.* **425**, 1097. DOI.

Robbrecht, E., Berghmans, D., Van der Linden, R.: 2009, Automated LASCO CME catalog for solar cycle 23: are CMEs scale invariant? *Astrophys. J.* **691**, 1222. DOI.

Rouillard, A.P., Davies, J.A., Forsyth, R.J., Rees, A., Davis, C.J., Harrison, R.A., Lockwood, M., Bewsher, D., Crothers, S.R., Eyles, C.J., Hapgood, M., Perry, C.H.: 2008, First imaging of corotating interaction regions using the STEREO spacecraft. *Geophys. Res. Lett.* **35**, L10110. DOI. ADS.

Sheeley, N.R. Jr., Herbst, A.D., Palatchi, C.A., Wang, Y.-M., Howard, R.A., Moses, J.D., Vourlidas, A., Newmark, J.S., Socker, D.G., Plunkett, S.P., Korendyke, C.M., Burlaga, L.F., Davila, J.M., Thompson, W.T., St Cyr, O.C., Harrison, R.A., Davis, C.J., Eyles, C.J., Halain, J.P., Wang, D., Rich, N.B., Battams, K., Esfandiari, E., Stenborg, G.: 2008, Heliospheric images of the solar wind at Earth. *Astrophys. J.* **675**, 853. DOI. ADS.

Sheeley, N.R., Walters, J.H., Wang, Y.-M., Howard, R.A.: 1999, Continuous tracking of coronal outflows: two kinds of coronal mass ejections. *J. Geophys. Res.* **104**, 24739. DOI. ADS.

Tappin, S.J.: 2017, Considerations for the use of STEREO-HI data for astronomical studies. *Astron. J.* **153**, 164. DOI.

Tappin, S.J., Howard, T.A., Hampson, M.M., Thompson, R.N., Burns, C.E.: 2012, On the autonomous detection of coronal mass ejections in heliospheric imager data. *J. Geophys. Res.* **117**, A05103. DOI. ADS.

Vourlidas, A., Balmaceda, L.A., Stenborg, G., Dal Lago, A.: 2017, Multi-viewpoint coronal mass ejection catalog based on STEREO COR2 observations. *Astrophys. J.* **838**, 141. DOI. ADS.

Wang, Y.-M., Colaninno, R.: 2014, Is solar cycle 24 producing more coronal mass ejections than cycle 23? *Astrophys. J. Lett.* **784**, L27. DOI. ADS.

Webb, D.F., Howard, T.A.: 2012, Coronal mass ejections: observations. *Living Rev. Solar Phys.* **9**, 3. DOI.

Webb, D.F., Jackson, B.V.: 1990, The identification and characteristics of solar mass ejections observed in the heliosphere by the HELIOS 2 photometers. *J. Geophys. Res.* **95**, 20641. DOI. ADS.

Yashiro, S., Michalek, G., Gopalswamy, N.: 2008, A comparison of coronal mass ejections identified by manual and automatic methods. *Ann. Geophys.* **26**, 3103. DOI. ADS.

Yashiro, S., Gopalswamy, N., Michalek, G., St. Cyr, O.C., Plunkett, S.P., Rich, N.B., Howard, R.A.: 2004, A catalog of white light coronal mass ejections observed by the SOHO spacecraft. *J. Geophys. Res.* **109**, A07105. DOI. ADS.

Zuccarello, F.P., Bemporad, A., Jacobs, C., Mierla, M., Poedts, S., Zuccarello, F.: 2011, The role of streamers in the deflection of coronal mass ejections: comparison between STEREO three-dimensional reconstructions and numerical simulations. *Astrophys. J.* **744**, 66. DOI.

Authors and Affiliations

L. Rodriguez[1] · **D. Barnes**[2] · **S. Hosteaux**[3] · **J.A. Davies**[2] · **S. Willems**[1] · **V. Pant**[4] · **R.A. Harrison**[2] · **D. Berghmans**[1] · **V. Bothmer**[5] · **J.P. Eastwood**[6] ·

P.T. Gallagher[7,8] ⓘ · **E.K.J. Kilpua**[9] ⓘ · **J. Magdalenic**[1,3] ⓘ · **M. Mierla**[1,10] ⓘ · **C. Möstl**[11] ⓘ ·
A.P. Rouillard[12] ⓘ · **D. Odstrčil**[13] ⓘ · **S. Poedts**[3,14] ⓘ

✉ L. Rodriguez
luciano.rodriguez@observatory.be

1   Solar-Terrestrial Centre of Excellence – SIDC, Royal Observatory of Belgium, Avenue
    Circulaire 3, 1180 Brussels, Belgium

2   STFC RAL Space, Rutherford Appleton Laboratory, Harwell Campus, Oxfordshire, OX11 0QX,
    UK

3   CmPA/Department of Mathematics, KU Leuven, Celestijnenlaan 200 B, 3001 Leuven, Belgium

4   Aryabhatta Research Institute of Observational Sciences, Nainital, 263001, India

5   Institute for Astrophysics, University of Göttingen, 37077 Göttingen, Germany

6   Blackett Laboratory, Imperial College London, SW7 2AZ, UK

7   School of Physics, Trinity College Dublin, Dublin 2, Ireland

8   School of Cosmic Physics, Dublin Institute for Advanced Studies, Dublin 2, Ireland

9   Department of Physics, University of Helsinki, PO Box 64, 00014, Helsinki, Finland

10 Institute of Geodynamics of the Romanian Academy, Bucharest, Romania

11 Space Research Institute, Austrian Academy of Sciences, 8042 Graz, Austria

12 Institut de Recherche en Astrophysique et Planétologie, 31028 Toulouse Cedex 4, France

13 School of Physics, Astronomy and Computational Sciences, George Mason University, Fairfax, VA
    22030-4444, USA

14 Institute of Physics, University of Maria Curie-Skłodowska, 20-031 Lublin, Poland