The Astrophysical Journal, 833:80 (15pp), 2016 December 10 doi:10.3847/1538-4357/833/1/80

AUTOMATED DETECTION OF CORONAL MASS EJECTIONS IN *STEREO* HELIOSPHERIC IMAGER DATA

V. Pant[1], S. Willems[2], L. Rodriguez[2], M. Mierla[2,3], D. Banerjee[1,4], and J. A. Davies[5]
[1] Indian Institute of Astrophysics, Bangalore-560 034, India
[2] Solar-Terrestrial Center of Excellence, Royal Observatory of Belgium, Avenue Circulaire 3, B-1180 Brussels, Belgium
[3] Institute of Geodynamics of the Romanian Academy, Bucharest, Romania
[4] Center of Excellence in Space Sciences, IISER Kolkata, India
[5] RAL Space, STFC Rutherford Appleton Laboratory, Harwell Campus, OX11 OQX, UK
Received 2016 April 1; revised 2016 October 4; accepted 2016 October 5; published 2016 December 8

ABSTRACT

We have performed, for the first time, the successful automated detection of coronal mass ejections (CMEs) in data from the inner heliospheric imager (HI-1) cameras on the *STEREO-A* spacecraft. Detection of CMEs is done in time–height maps based on the application of the Hough transform, using a modified version of the CACTus software package, conventionally applied to coronagraph data. In this paper, we describe the method of detection. We present the results of the application of the technique to a few CMEs, which are well detected in the HI-1 imagery, and compare these results with those based on manual-cataloging methodologies. We discuss, in detail, the advantages and disadvantages of this method.

Key words: Sun: corona – Sun: coronal mass ejections (CMEs) – Sun: heliosphere

Supporting material: animations

1. INTRODUCTION

According to the original definition, coronal mass ejections (CMEs) are observable changes in coronal structure, which occur on timescales from a few minutes to several hours and involve the appearance (Hundhausen et al. 1984) and outward propagation (Schwenn 1996) of new, discrete and bright white-light features in the coronagraph field of view. CMEs result from the episodic expulsion of plasma and magnetic field from the solar atmosphere into the heliosphere with speeds, which are typically 400 km s^{-1}, but which can range from 100–2500 km s^{-1} (Yashiro et al. 2004; Manoharan & Mujiber Rahman 2011). CMEs are considered the most energetic events in the solar system. Furthermore, they are very important in terms of space weather, being the drivers of the largest geomagnetic storms detected on Earth. Since 1996, we have been able to monitor CMEs routinely from the L1 vantage point using the Large Angle Spectrometric Coronagraph (LASCO) (Brueckner et al. 1995) on the *Solar and Heliospheric Observatory* (*SOHO*) spacecraft. Moreover, since late 2006, we have also been able to monitor CMEs from a location off the Sun–Earth line using the COR1 and COR2 coronagraphs, which form part of the Sun Earth Connection Coronal and Heliospheric Investigation (SECCHI) (Howard et al. 2008) imaging package on the twin-spacecraft *Solar Terrestrial Relation Observatory* (*STEREO*) mission. The SECCHI Heliospheric Imager (HI) instruments on *STEREO* effectively extend the coronagraph observing methodology to larger distances from the Sun by providing 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°, with the FOVs being centered, in nominal operations, at around 14° and 54° elongation, respectively (Socker et al. 2000). The extensive HI FOV allows us to observe CMEs propagating over vast distances of interplanetary space (Eyles et al. 2007; Harrison et al. 2008, 2009; Davis et al. 2009). The concept of wide-angle heliospheric imaging was first demonstrated by the Solar Mass Ejection Imager (SMEI) on board the Earth-orbiting *Coriolis* spacecraft (Eyles et al. 2003). The terminology of interplanetary CMEs is often applied to the interplanetary counterparts of CMEs (see, Gopalswamy et al. 2000; Riley et al. 2006; Zurbuchen & Richardson 2006). However, due to the success of *Coriolis*/SMEI and *STEREO*/HI in filling the vast observational gap between coronagraph imagery and in situ measurements, some authors are now suggesting that the CME terminology should be applied to both phenomena (Webb & Howard 2012; Bisi et al. 2013).

Since the launch of *STEREO*, CMEs have been identified through visual inspection of HI images and the resultant event catalogs have been made public (Barnard et al. 2014, 2015; R. A. Harrison et al. 2016, in preparation). However, such visual identification of CMEs is biased by human subjectivity and hence CME detection may or may not be consistent over an extended period of time (Wang & Colaninno 2014). It is worth mentioning that each CME in the Solar Stormwatch CME catalog (Barnard et al. 2014) is identified manually by multiple independent operators in order to reduce the subjectivity of the detection. The properties of a given CME are derived by averaging the independent detections. Nevertheless, automated detection offers the capability of providing more objective CME detection. One such software package, Computer Aided CME Tracking (CACTus) was developed to detect CMEs in LASCO/C2 and C3 coronagraph images (Robbrecht & Berghmans 2004) and subsequently applied to COR2 coronagraph imagers from *STEREO*. CACTus applies the original definition of a CME; "a CME is a new, discrete, bright, white-light feature in the coronagraph FOV with a radially outward velocity" (Robbrecht & Berghmans 2004). CME detection using CACTus is more objective and faster compared with the visual identification, because CMEs are detected, and characterized according to a strict set of precisely defined constraints. The catalogs produced using CACTus (available online[6]) are similar to manually compiled

[6] http://sidc.oma.be/cactus

catalogs in terms of the parameters they contain. Robbrecht et al. (2009) have compared the CME parameters derived by CACTus with those obtained by manual detection of CMEs in LASCO/C2 and C3 images. Apart from CACTus, there are several other catalogs of CMEs, which have been identified automatically in coronagraph data. The Coronal Image Processing (CORIMP) automated detection algorithm uses normalized radial gradient filtering and deconvolution to separate quiescent structures (background corona) and dynamic structures (features such as CMEs, which propagate radially outward) (Byrne et al. 2012; Morgan et al. 2012). CORIMP has been used to detect CMEs automatically in LASCO and COR2 images. Another such algorithm, the Automatic Recognition of Transient Events and Marseille Inventory from Synoptic maps (ARTEMIS), was developed by Boursier et al. (2009). ARTEMIS detects CMEs automatically in LASCO synoptic maps using image filtering and segmentation techniques. Furthermore, the Solar Eruptive Events Detection System (SEEDS) also detects CMEs in polar transformed running difference images from LASCO and COR2 (Olmedo et al. 2008). SEEDS isolates the leading edges of CMEs by intensity thresholding; by tracking CMEs in sequential images, their speed and acceleration are calculated. Byrne (2015) compared the CORIMP catalogs with catalogs generated using other automated (i.e., CACTus and SEEDS) and manual detection methods (i.e., the CDAW catalog). The authors demonstrated the robustness of CORIMP in deriving the kinematics of the automatically detected CMEs. The automated detection of CMEs in the heliosphere is, however, not an easy task, mainly due to their low brightness compared to that of the other contributions to the white-light signal (principally the F-coronal and stellar backgrounds). Despite this, there have been previous attempts to automatically identify CMEs in heliospheric images. The Automated Interplanetary Coronal Mass Ejection Detection (AICMED) tool was developed by Tappin et al. (2012) to detect CMEs in SMEI data, in particular. Like CACTus, AICMED works on the principle of the Hough transform, which detects straight ridges. AICMED uses the Hough transform to detect curved ridges in time-elongation maps (commonly called J-maps) by splitting each curved ridge into several straight ridges (this curvature, as will be discussed later, is a geometric artifact associated with imaging out to large elongations). However, due to a number of features in the SMEI data (such as rings generated by hot pixels, scattered light from the moon, cometary tails, high-altitude aurora and particle hits during crossings of the South Atlantic Anomaly (SAA) and auroral zones), there were many false detections. Recently, Barnard et al. (2015) reported the differences in velocities estimated using automated and manual tracking of ridges in time-elongation maps created from *STEREO*/HI-1 and HI-2 images (see also, Savani et al. 2009), although the authors did not perform a completely automatic detection of CMEs in the HI imagery.

We have adapted the CACTus software package to automatically detect CMEs in *STEREO*/HI-1 data. In this manuscript, we first describe the methodology, which we have used (Section 2). Subsequently, in Sections 3 and 4, we compare the automatically derived parameters for a selection of CMEs with analogous entries in manual catalogs for a selection of CMEs: namely the time of appearance (t_0), central position angle (CPA) of propagation, position angle (PA), width (da), and the projected speed (v).

2. METHOD OF DETECTION

In this section, we describe the method of automated detection of CMEs in HI-1 data from *STEREO-A*, which is an adaptation of the original CACTus methodology, as discussed in Robbrecht & Berghmans (2004), but with some modifications in order to make it work with the HI images. CME detection by CACTus is based on the principle of the Hough transform (see, Jahne 1997), which can be used to detect straight lines in noisy data. The brightness of CMEs in heliospheric images is generally lower relative to other contributions to the signal (such as F-coronal and stellar backgrounds) than in coronagraph images, leading to an inferior signal-to-noise ratio. Since the automated detection of CMEs depends critically on their clarity, we need to carefully process the images before implementing the Hough transform to the time–height maps.

2.1. Preprocessing of HI Images

1. *Data acquisition:* The 1 day background-subtracted level 2 HI-1 science images (array size, 1024 × 1024) in units of DN s^{-1} per CCD pixel are downloaded via the UK Solar System Data Center (UKSSDC) website[7] (note that pixels on the CCD detector are binned 2 × 2 onboard to generate the science images; our subsequent use of the term pixel refers to a pixel in a science image). These images have a nominal cadence of 40 minutes. For illustration throughout Section 2, we analyze HI-1 images from the period extending from the beginning of 2010 April 02 to the end of 2010 April 04, encompassing about 72 hr. However, in Section 3, when comparing with the results of manual cataloging, we include five additional days in our analysis. The background, which is subtracted from each HI-1 image, is the average of the lowest quartile of the data in each pixel within the FOV over a 1 day period centered on the image of interest[8] (Tappin et al. 2016). By subtracting a daily background, quasistatic components of the signal, principally the F-corona and also the less variable elements of the K-corona (e.g., streamers) are removed. Such a 1 day background-subtracted level 2 HI-1 image from *STEREO-A* is shown in the left-hand panel of Figure 1. The processing, which has been applied to the images, is described on the UKSSDC website.[9]
2. *Removal of stars and planets:* It is evident from this image that the white-light signal due to the presence of bright stars and planets in the FOV can exceed the CME signal (the leading edge of the CME is marked with a yellow arrow). The stars and planets are not removed by the background-subtraction procedure as they move through the FOV at a relatively fast rate. To reduce the effects of the stars and the planets we use a sigma filter, which works by first computing the mean and the standard deviation of the intensity of the neighboring pixels excluding the pixel on which it is centered. If the intensity of the central pixel is greater than a chosen threshold (mean + 4 × standard deviation), its value is

[7] http://www.ukssdc.rl.ac.uk/solar/stereo/data.html
[8] http://www.ukssdc.rl.ac.uk/solar/stereo/documentation/HI_processing_L2_data.html
[9] http://www.ukssdc.rl.ac.uk/solar/stereo/documentation/HI_processing.html

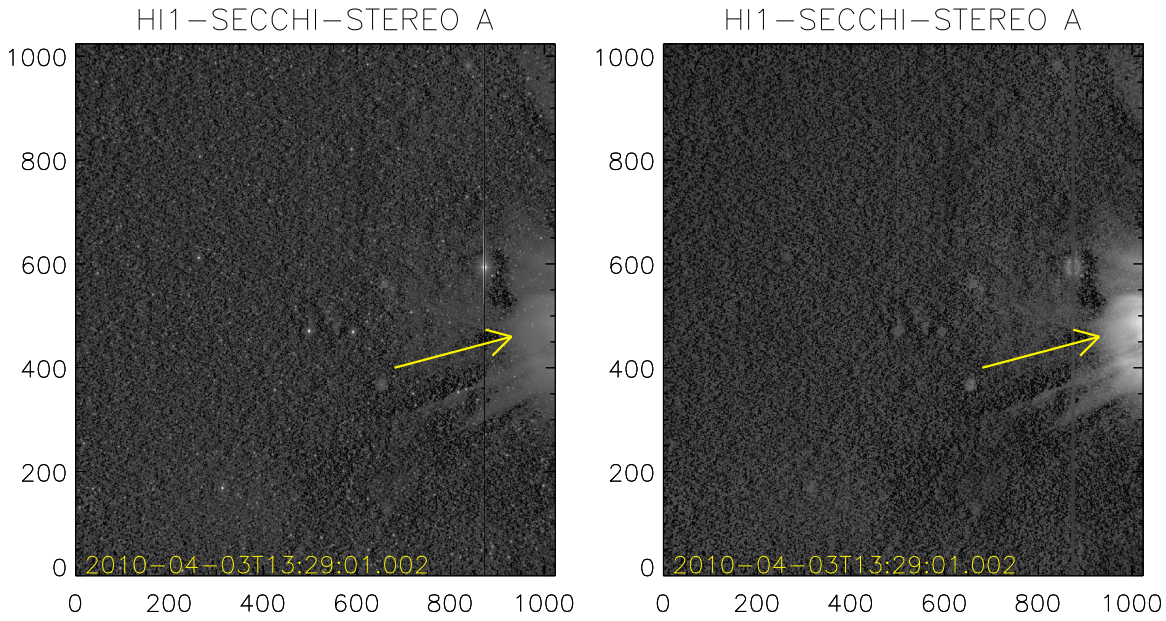

Figure 1. Left: level 2 *STEREO-A* HI-1 image after 1 day background subtraction and the initial removal of bright streaks (most clearly that resulting from the presence of Mercury in the FOV). Right: as left, but with further processing to reduce the effects of bright planets and stars (as discussed in the text). The yellow arrow indicates the leading edge of a CME (at a PA of 102°) observed on 2010 April 03.

replaced by the mean value of its neighboring pixels. This process is iterated over all of the pixels in an image, recursively up to 20 times or until no further change is observed. By examining a few test cases, we determined that using a sigma filtering with a width of 50 pixels works well at reducing the effects of planets and stars.

3. *Removal of bright streaks associated with planets and bright stars:* Unlike for coronagraphs with their smaller FOVs, in which only a few planets and bright stars tend to be present at one time, more planets and bright stars are present in the somewhat larger HI-1 FOV. Planets and bright stars are associated with bright vertically extended streaks, which result from vertical blooming of the signal in saturated pixels. It is crucial to remove these bright vertical streaks prior to CME detection. It should be noted that the *secchi_prep.pro* routine (which is applied to the HI data as part of the processing performed at the UKSSDC) attempts to replace the bright vertically extended streaks with NaN values (NaN streak). However, we find that this procedure is often not always totally effective. An incompletely removed bright streak may still be brighter than a CME in the difference images, which are used for CME detection by CACTus (see Section 2.2) and, therefore, may affect the CME detection. Hence, CACTus performs additional processing of the HI-1 data, which is not present in earlier versions of the software, in order to remove these residual bright streaks. This process involves the use of dilation. Dilation is an image processing technique, which "grows" or "widens" an object. The extent of the widening is based on the shape of the kernel (Gonzalez & Woods 2002). We choose a 1D kernel with a horizontal width of 6 pixels for dilation. This width of kernel is found to work well at covering any residual bright vertical streaks associated with a planet or bright star. Upon its application, the original NaN streak, which is implemented in the UKSSDC processing, becomes wider, as a result of which any residual bright streak is completely replaced by NaNs. This process results in the highly effective replacement of the residual bright streaks in the HI-1 images. Subsequently, we use convolution based on a 1D kernel of 100 pixels in width to replace the NaN streak with values of the neighboring (horizontal) pixels (note that the NaN streak extends over the entire vertical range of the image). Through this procedure, we isolate the position of the dilated NaN streak and replace it with the intensity obtained through convolution. This fills the affected pixels with the values of the surrounding pixels without smoothing the entire image.

The processed image after the application of the above procedures is shown in the right-hand panel of Figure 1. The CME is now brighter relative to the stars and planets, which are suppressed and the residual bright streak (with its associated NaN streak) is now filled with neighboring intensity values.

2.2. Polar Transformation

The next step toward the detection of CMEs in *STEREO*/HI imagery is the conversion of the helioprojective Cartesian (HPC) coordinates provided as one of the standard coordinate system of the HI images, into polar ($\theta - r$) coordinates, where θ is the PA, measured counterclockwise from solar north, and r is the distance from the Sun-center projected onto the plane of sky (POS), as seen from the spacecraft.

1. In order to do this, we first convert the HPC coordinates of the HI images to helioprojective-radial (HPR) coordinates (see Thompson 2006) using routines, which are available in the SolarSoft package. It should be noted that the accurate pointing calibration of the HI imagery is achieved in the UKSSDC preprocessing through comparison of the star field within each image with a star catalog (Brown et al. 2009).
2. The conversion to HPR, assigns each pixel in the HI image with two values, the PA (θ) and the elongation (ϵ), the angle between its line of sight (LOS) and the

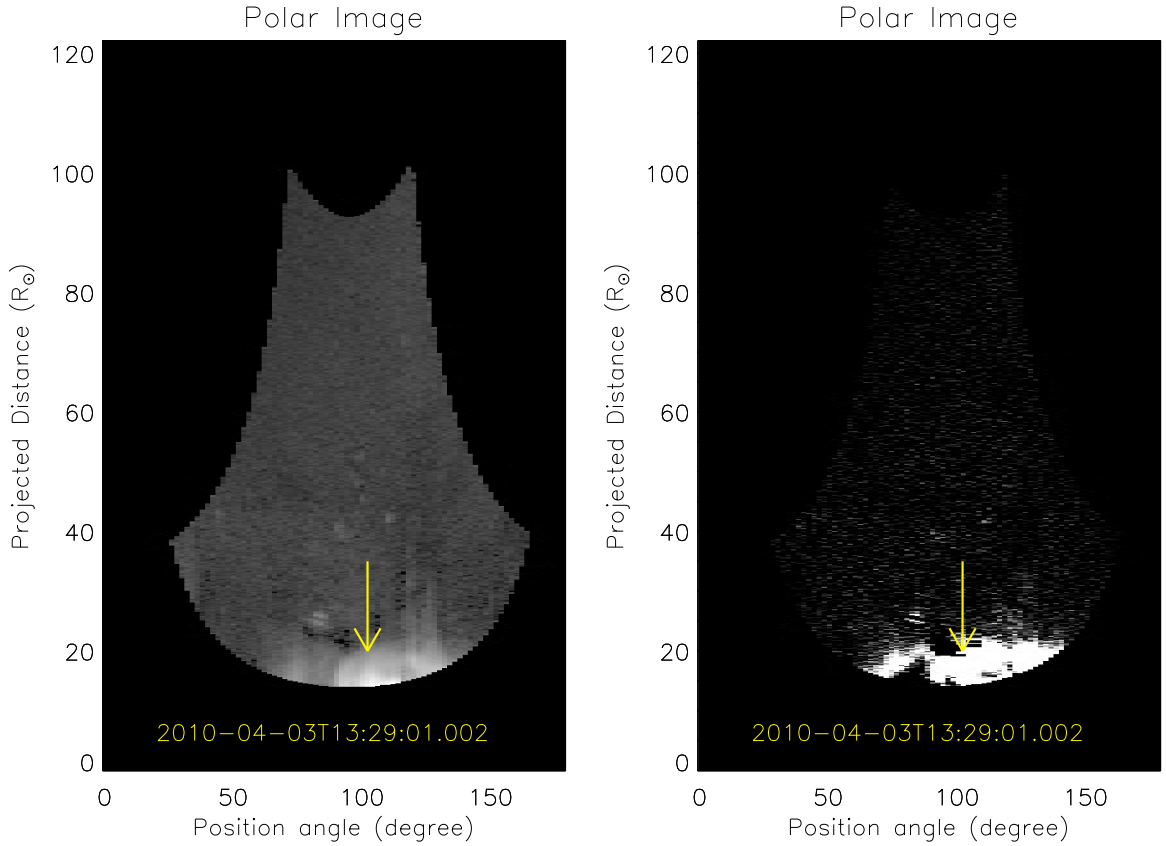

Figure 2. Left: polar transformation of the image presented in the right-hand panel of Figure 1. Right: difference and thresholded version of the left-hand image. The yellow arrows represent the leading edge of the CME co-spatial with the arrow shown in Figure 1.

spacecraft-Sun line. Knowing the elongation of each pixel in the image, we can calculate its projected distance from the Sun-center on the POS, using the expression

$$r = d\,\tan(\epsilon), \tag{1}$$

where d is the distance of the observing spacecraft from the Sun.

3. The polar ($\theta - r$) representation of the image presented in the right-hand panel of Figure 1 is shown in the left-hand panel of Figure 2, where the x-axis represents PA (θ) and the y-axis represents the projected distance in the POS derived using Equation (1). The data are binned such that one height bin corresponds to 100,000 km (roughly 1/7th of solar radius). Since CMEs are large-scale features, this binning does not lose any structural information. Furthermore, binning reduces noise and computation time.
4. To remove quasistatic features, which vary on a timescale shorter than the 1 day subtracted background, and longer than the 40 minute image cadence of HI-1, running difference $\theta - r$ images are generated automatically in a manner analogous to what is done in the conventional version of CACTus (Robbrecht & Berghmans 2004). In particular, this removes much of the shorter timescale variation of the streamer belt. The right-hand panel of Figure 2 shows only the positive values of the difference image (corresponding to an increase in brightness, hence density, relative to the previous image).

2.2.1. Geometric Models and Fitting Methods

Since the advent of wide-angle imaging of the solar wind, by *Coriolis*/SMEI and *STEREO*/HI, a number of geometries have been derived, which enable the elongation angle of a CME (or other solar wind structure) to be converted to a radial distance from the Sun-center, namely Point-P (PP) (Howard et al. 2006), Fixed-ϕ (FP) (Kahler & Webb 2007; Rouillard et al. 2008; Sheeley et al. 2008a), Harmonic Mean (HM) (Lugaz et al. 2009), and Self-Similar Expansion (SSE) (Davies et al. 2012). The FP, HM, and SSE conversion methods require knowledge, or assumption, of the 3D propagation direction of a solar wind transient (such as from its source location on the Sun). As we do not know the propagation direction of the CME a priori, it is easiest to assume that it propagates at an angle of 90° from the Sun-spacecraft line (i.e., in the POS). The POS approximation (as described by Equation (1)) is, in fact, a special case of the FP conversion methodology. The PP approximation works on the principle that the maximum brightness contribution along any LOS comes from the point closest to the Sun (i.e., the Thomson sphere), as discussed by, for example, Vourlidas & Howard (2006) and Howard & DeForest (2012). Howard & DeForest (2012) and Inhester (2015) have shown that, in fact, Thomson scattering maximizes over a wide range of angles around the Thomson sphere (the so-called Thomson plateau). We discount the use of PP, mainly because it traces out a non-radially propagating point on a CME front. Here, as discussed in Section 2.2, we adopt the POS approximation for deriving radial distance from elongation, which is consistent with the original CACTus ethos.

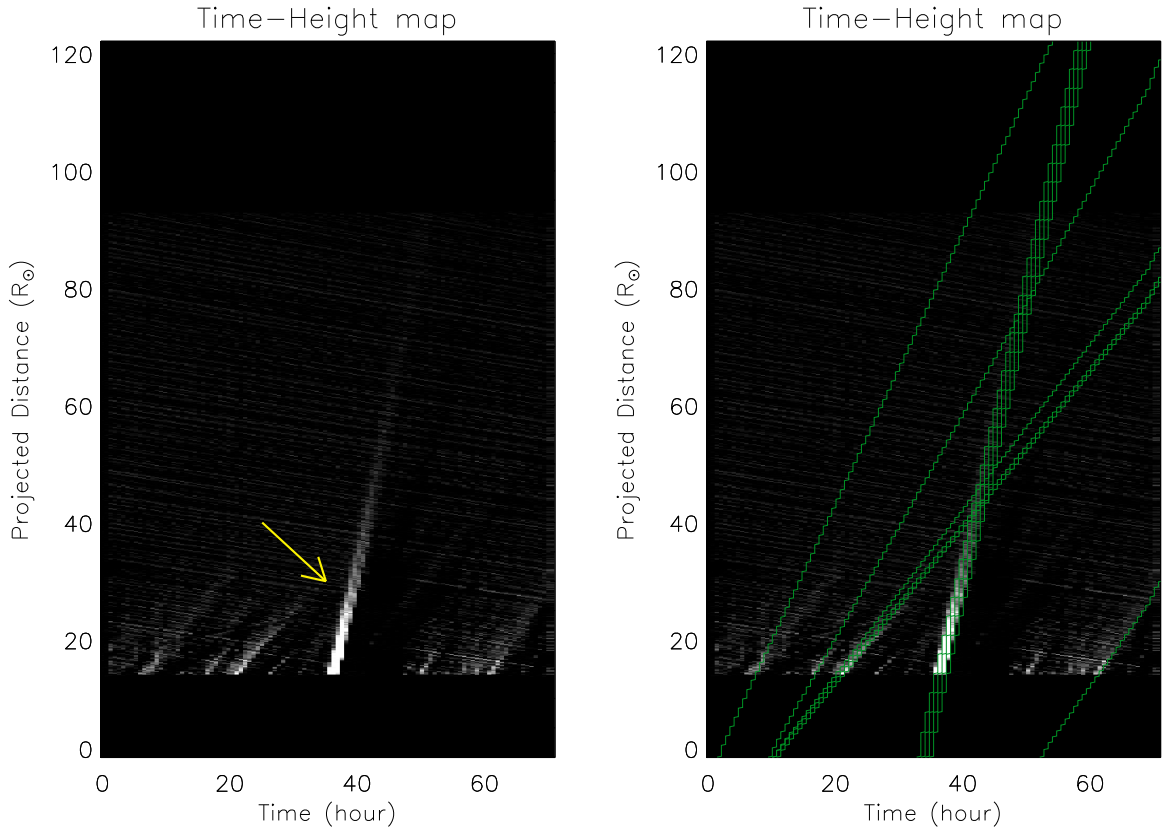

Figure 3. Left: time–height map at 102° PA, generated from HI-1 images from *STEREO-A*, covering the interval that extends from 2010 April 02 to 04. The yellow arrow indicates the inclined ridge corresponding to the leading edge of the particular CME of interest. Right: as for the left, but overplotted with green curves, which represent the significant ridges detected through the application of the Hough transform.

It is worth making the point here that it is actually possible to estimate the 3D propagation angle (and radial speed) of a solar wind feature, such as a CME, without having a priori knowledge of its source region, provided that it can be tracked over sufficient elongation extent. This can be done by analyzing the curvature of its associated ridge in the J-map. Such curvature is present even if the feature is propagating at a constant speed in a fixed direction (see, for example, Figure 2 of Davies et al. 2012, where the authors plot simulated time-elongation profile curves for FP, HM, and SSE geometries). By assuming a fixed geometry, the time-elongation profile of the ridge can be analyzed to provide an estimate of the 3D propagation direction and radial speed. The current version of CACTus can only detect straight lines. Hence, we cannot estimate the 3D propagation angle in this way. Möstl et al. (2014) have reported that CMEs can decelerate out to 1 AU. Thus, in addition to the geometric artifact discussed above, any deceleration (or indeed acceleration) of a CME during its propagation will modify the curvature due to aforementioned effect. However, it should be noted that, even if some curvature is present, CACTus can still detect a ridge if it is sufficiently bright (such as the ridge detected at 20 hr at 15 $R_\odot$ in Figure 3). However, any velocity estimate may not be accurate.

We do not perform CME detection in HI-2 imagery, not least because of the inherent curvature in the signature of a CME propagating over a large range of elongations (Davies et al. 2012). As noted above, such curvature could adversely affect CME detection, when using the Hough transform implemented in CACTus, as the Hough transform only detects straight lines. However, in the case of observations over a limited elongation extent (less than some 15°), the track of a CME propagating radially outwards at a constant velocity would be virtually straight. In practice, there would still be a slight curvature over the elongation range covered by HI-1, but not enough to affect the performance of the algorithm, particularly since CMEs are mainly detected from the inner edge of the HI-1 FOV, an elongation of 4° out to 15°; the latter equated to a POS distance of 57 $R_\odot$. The slightly curved ridge in the left-hand panel of Figure 3 (detected in HI-1 FOV at 20 hr at 15 $R_\odot$) is successfully detected by Hough transform (shown in the right-hand panel of Figure 3), because it is bright. However, the speed estimate in this and similar cases may not be accurate.

Difference $\theta - r$ images (similar to the one that is shown in the right-hand panel of Figure 2) are stacked over time to produce a $[\theta, r, t]$ data cube, where θ is the PA, r is the radial distance and t is the time in hours from the start of the interval under analysis (up to 72 hr in this case).

2.3. Application of the Hough Transform

1. For each value of the PA, θ, in the resultant $[\theta, r, t]$ data cube, we have what is effectively a time–height ($[t, r]$) map. The left-hand panel of Figure 3 shows the time–height map at $\theta = 102°$, for the interval covering the propagation of the CME shown in Figures 1 and 2 through the FOV of HI-1 on *STEREO-A*. This PA corresponds to the the yellow arrows in Figures 1 and 2. The x-axis represents the time in hours from 00:00 UT on 2010 April 02 and the y-axis represents the POS distance from the Sun-center.

2. Inclined ridges with a positive slope represent features traveling away from the Sun, such as the leading edge of the CME shown in Figures 1 and 2 (the track of this CME front is indicated by a yellow arrow in Figure 3). Faint ridges with negative slope correspond to the residual star signal. Prior to superior conjunction (since which the *STEREO-A* spacecraft has been rotated by 180°), the motion of *STEREO-A* was such that the apparent drift of the star field across HI images was antisunward (i.e., in the opposite direction to CMEs). One should be aware that there could be multiple CMEs propagating along a particular PA over any extended time range (which is why we mark the ridge corresponding to the CME of interest here with a yellow arrow).
3. The Hough transform is defined as the mapping from image space (in this case a time–height map) to parameter space (or accumulator space) (see, Jahne 1997; Robbrecht & Berghmans 2004). An inclined ridge is uniquely characterized by two parameters, its slope (m) and intercept (c), provided the slope is not infinite. Since CMEs propagate anti-antisunward with finite velocities, the inclined ridges related thereto have positive finite slope. One ridge maps to a single point in accumulator space (see Figure 2 in Robbrecht & Berghmans 2004). The weight given to the point in accumulator space is determined by the number of points lying along the ridge in the time–height map. Brighter ridges in the time–height map tend to be given larger weights in accumulator space than fainter ridges, as they can usually be tracked out further in radial distance.
4. As noted above, CACTus uses the Hough transform to isolate (as straight lines) the significant ridges in time–height maps associated with the passage of CMEs. We filter out those points in accumulator space with a low weighting by applying a weight threshold, $W_{\rm thresh,}$ which is given by,

$$W_{\rm thresh} = W_{\rm mean} + f \times W_{\rm sd} \quad (2)$$

where, $W_{\rm mean}$ and $W_{\rm sd}$ are the mean and standard deviation of the weights of all points in accumulator space and f is an arbitrary factor. CACTus uses an empirically based value of $f = 4$ in the analysis of the *STEREO*/HI-1 data, based on analyzing a sample of CMEs.
5. Points in accumulator space with weights, which exceed $W_{\rm thresh}$, are taken to be the points, which correspond to the tracks of CMEs in time–height maps. The time–height map presented in the right-hand panel of Figure 3 reproduces that shown in the left-hand panel, but with significant ridges (corresponding to those points in accumulator space, which exceed $W_{\rm thresh}$) overplotted in green. We plot the green curves over the full height range. The coordinates of the detected points in accumulator space, i.e., m and c of each detected ridge in the time–height map, provide an estimate of the POS velocity (v) and time of first appearance of each detected CME (t_0), respectively. t_0 does not represent the time that the CME first appears in the HI-1 FOV but, instead, corresponds to the back-projected time at which the CME would be at the Sun-center (i.e., $r = 0$). This is because the data cube input into CACTus extends from $r = 0$. To estimate the time at which the CME enters the HI-1 FOV, a correction to t_0 is required; no such correction to the velocity is required. The correction depends on the velocity of a ridge and the distance to the inner edge of the HI-1 FOV. Unlike coronagraph images, where the inner edge of the FOV is independent of θ (because the occulter is circular), the POS distance of the inner edge of the HI-1 FOV depends on θ (see Figure 2). For the time–height map presented in Figure 3, this distance is 15 $R_\odot$. The θ-dependent correction of t_0, which is implemented in the current version of CACTus, was not present in earlier versions. As discussed above, a ridge in time–height map corresponds to the propagation of a CME along a particular PA (θ). We perform the Hough transform on time–height maps at every PA within the HI-1 FOV. Since CMEs are extended in θ, a single CME will correspond to a cluster of points in the resultant [θ, v, t_0] data cube.
6. CACTus subsequently performs integration over v in order to obtain a [θ, t_0] map. We use a morphological closing technique to fill the gaps between the points of a cluster in such a map. This technique performs dilation of an image (as described above) followed by the contraction, using a kernel of a given size (see, Gonzalez & Woods 2002). This process fills gaps between the points in a cluster in the [θ,t_0] map, which are smaller than the size of the kernel.
7. Unlike the version of CACTus implemented on coronagraph images, which uses a 1D kernel with a width of 5 bins in the θ direction, a 2D kernel is used for detecting CMEs in HI-1 data. This kernel has a width of 8 bins in θ and 5 bins in t_0. This means that CACTus bridges the gaps between the points in accumulator space, which are less than 16° in θ and 200 minutes in t_0. These values are empirically chosen (through examination of a number of CMEs observed by HI-1) such that they connect different parts of the same CME together without (too often) combining different CMEs. Such morphological closing is required because, for example, a single CME often comprises more than one distinct feature (such as a leading edge and prominence material) and, moreover, the brightness of even a single feature may be nonuniform in θ.
8. The left panel of Figure 4 shows a [θ,t_0] map, which we will henceforth refer to as a CME map, which spans the same interval as presented in Figure 3 (i.e., 2010 April 02 to 04). The x-axis represents PA and y-axis, time. The CME map illustrated in the right-hand panel of Figure 4 presents the location, as a function of both PA and time, of the maximum POS velocity for each of the five CMEs detected by CACTus during the time interval covered by the time–height maps presented in Figure 3. The magnitude of the velocity is color coded according to the color bar. The white contours overplotted on the right-hand panel of Figure 4 illustrate the identified boundary of each cluster. Each cluster, of course, represents a separate CME detected by CACTus. Cluster 4 in the left-hand panel of Figure 4 corresponds to the CME under particular consideration here, images of which are shown in Figures 1 and 2.

It is also important to note that, for various reasons, there are occasional data gaps in the HI-1 telemetry stream. This not only leads to missing images, but also occasional incomplete

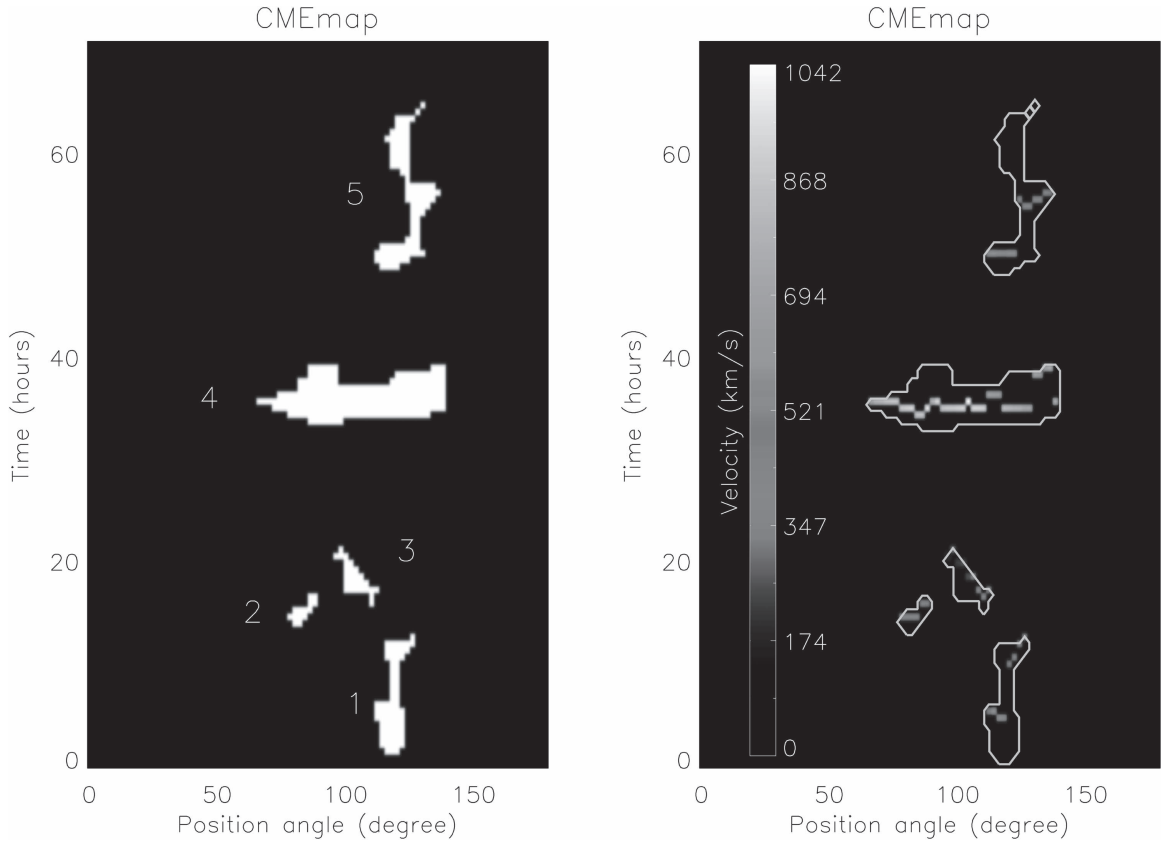

Figure 4. Left: CME map showing five clusters of points representing five different CMEs detected during the time interval extending from 2010 April 02 to 04. The clusters are numbered according to their time of first appearance. Right: CME map showing for each cluster (i.e., for each CME), the maximum velocity (color coded) at each PA over the entire PA extent of the CME. White contours represent the identified boundaries of the clusters.

images (so-called "missing blocks"). One advantage of the Hough transform is that it can detect ridges even in the presence of data gaps, provided the ridge is sufficiently bright and the data gap is sufficiently short (so that the weight of the point in Hough space corresponding to the ridge still exceeds $W_{\rm thresh}$). Since we apply morphological closing in θ and time, data gaps will not adversely affect the results, unless they are too long or too frequent. If there are too many data gaps, CACTus might not detect any ridges or may detect multiple ridges. While this could be resolved by resampling prior to implementing the Hough transform, there are too few data gaps in the HI-1 imagery for this to be necessary.

2.4. Determination of CME Position Angle (PA) Width, Time of Appearance and Velocity

1. *Estimation of PA width:* Once the so-called CME map has been created, the PA width of a CME is estimated by calculating the extent of the corresponding cluster in the CME map in the θ direction. It is important to note that for each CME the PA width provided by CACTus is the maximum width of that CME throughout its propagation through the HI-1 FOV. The CPA of propagation of a CME is calculated as the midpoint of its associated cluster in θ. Figure 5 shows difference images for two of the CMEs (CMEs that correspond to clusters 4 and 5 in Figure 4), which are detected by the application of CACTus to HI-1 imagery from *STEREO-A* during the time interval extending from 2010 April 02 to 04 (see animated Figures 7 and 8 available online). In each case, the PA width yielded by the method is delimited by white lines.
2. *Estimation of time of appearance:* To estimate the time of appearance of a specific CME in the HI-1 FOV, a local background intensity and standard deviation for the corresponding cluster in the CME map is first estimated using the $[\theta, r, t]$ data cube. For each cluster in a CME map, a local background intensity is estimated by fitting a straight line to the intensity, which is summed over the PA width and radial distance extent of the $[\theta, r, t]$ data cube over a time range, which extends from 30 images before the lower boundary of a cluster to 30 images after the upper boundary of the same cluster. The value of 30 is chosen empirically. The time of appearance of a CME is defined as being the time beyond/after the lower boundary of a cluster at which the intensity first exceeds the background value by 2σ, where σ is the standard deviation of the intensity. We do not simply use the lower boundary to determine the time of appearance as CMEs can generate waves or shocks, which can be manifest as regions of faintly enhanced brightness prior to the appearance of the CME itself. Moreover, small blobs are sometimes observed to propagate radially outward ahead of a CME (Robbrecht et al. 2009). Such brightenings ahead of CMEs may (or indeed may not) be detected by the Hough transform. If such features are identified in accumulator space, they may subsequently be amalgamated with the main cluster after the application of the morphological closing technique. Therefore, the lower boundary itself may not represent the true time

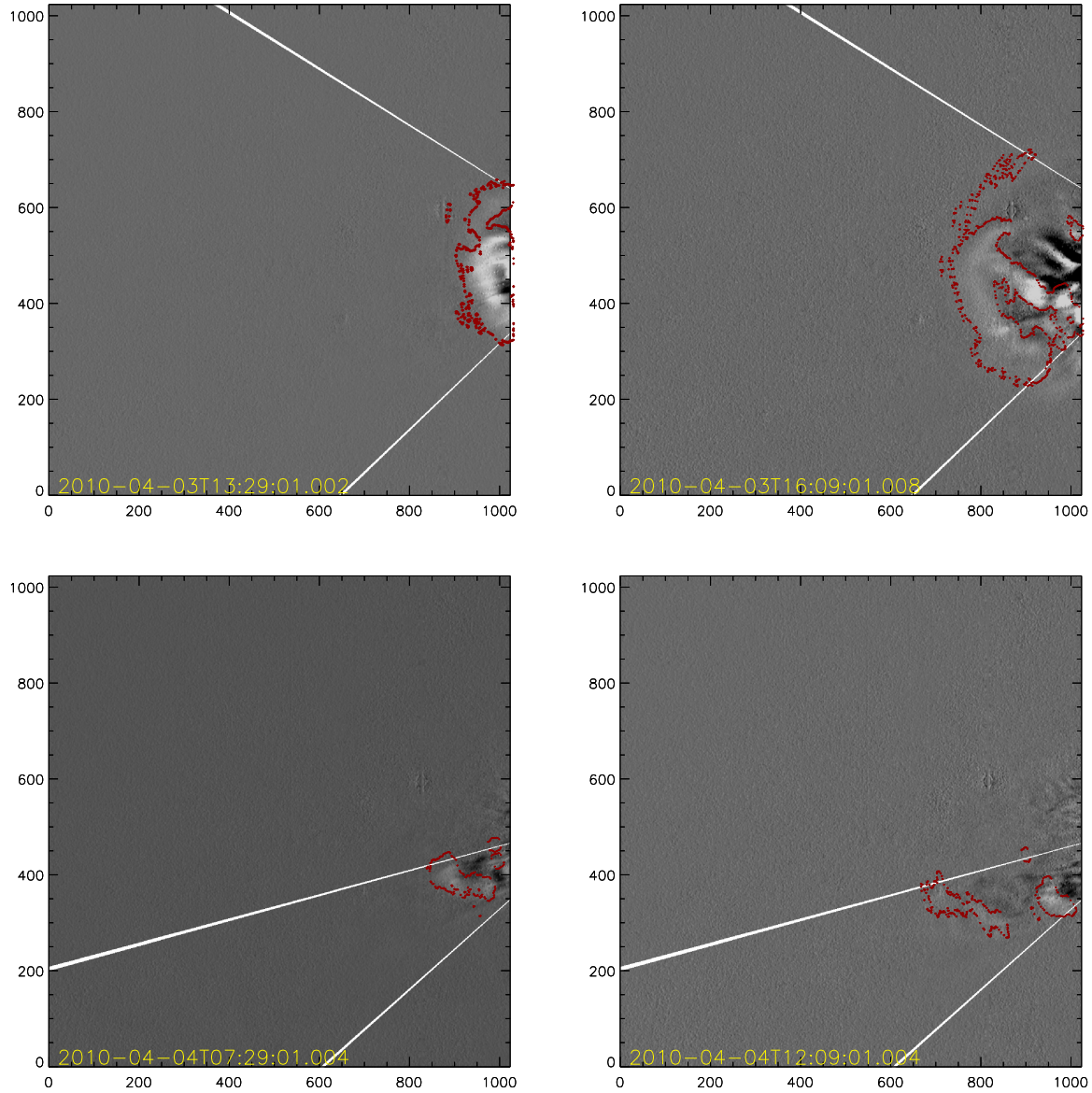

Figure 5. Top: difference images of the CME corresponding to cluster 4 in the left-hand panel of Figure 4 (No. 4 in Table 1). White lines delimit the northernmost and southernmost PA extents of the CME. Bottom: same as left-hand panel, but for cluster 5 (No. 5 in Table 1). Points overplotted in red represent the perimeter of the radially outward moving features identified by application of the Hough transform.

of CME appearance. If any such feature or unstructured flow (for example, streamer blobs) prior to CME appearance is bright enough to exceed the 2σ threshold, then it will affect the estimated time of arrival (potentially making it earlier than what would be deduced from visual inspection).

3. *Estimation of velocity:* Each bin in a CME map that falls within a CME cluster is associated with a velocity, v. For each cluster of points in a CME map, and thus for each CME, we identify the maximum velocity at each PA within the identified CME PA span (right-hand panel of Figure 4). The left-hand and right-hand panels of Figure 6 show the maximum velocity as a function of PA for the CMEs, which correspond to clusters 4 and 5, respectively, in Figure 4 (images of which are presented in Figure 5). Overplotted on each panel is a box-and-whisker key depicting the range of maximum velocities over the entire PA extent of the corresponding CME. The upper and lower boundaries of the box itself represent the upper and lower quartiles of the distribution of maximum velocities; the median value is marked by a horizontal line inside the box. The whiskers encompass the highest and the lowest value of the maximum velocity.

Estimated parameters of all five CMEs, which were detected by CACTus in HI-1 images from *STEREO-A*/HI-1 images during the time period extending from 2010 April 02 to 04 (including the two CMEs shown in Figure 5) are listed in Table 1.

3. COMPARISON WITH MANUAL CATALOG

We have applied CACTus to eight different days of *STEREO-A*/HI-1 images (2010 April 04, 2010 April 03, 2010 April 02, 2008 December 12, 2008 April 26, 2008 February 04, 2007 October 22, and 2007 April 19) and have compared the results with those of manual CME cataloging endeavors, in particular those performed as part of the EU FP7 Heliospheric Cataloging Analysis and Techniques Service

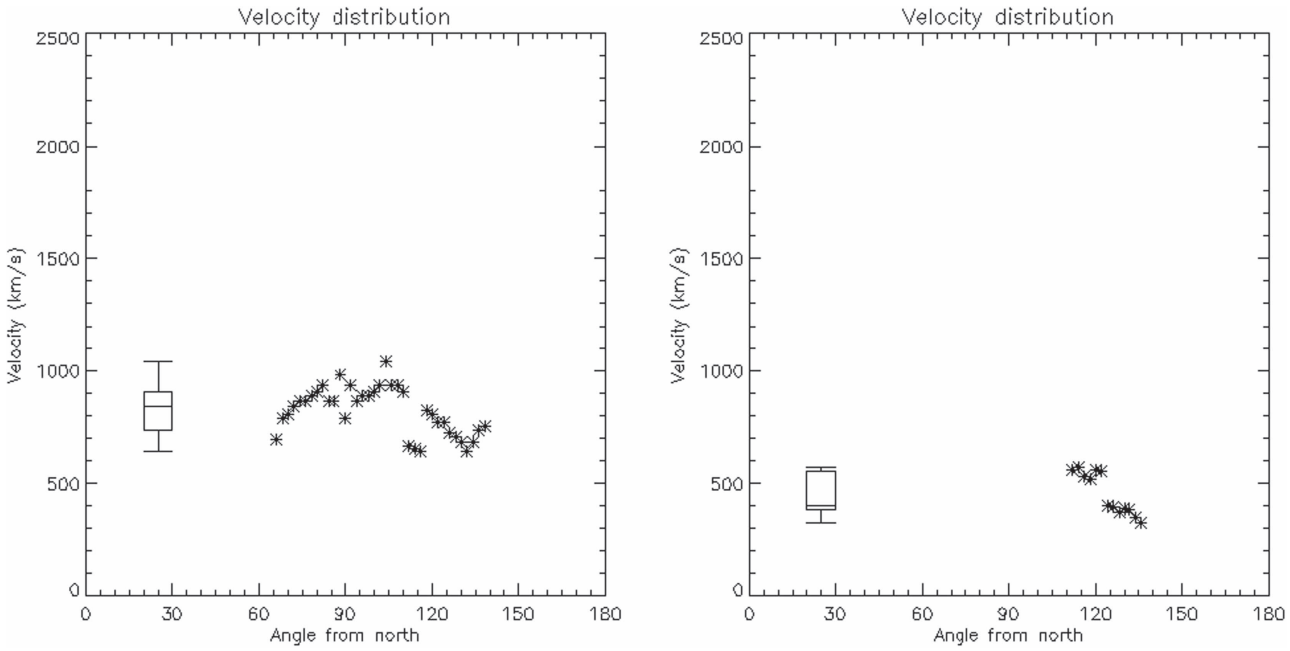

Figure 6. Left: maximum velocity as a function of PA (measured counterclockwise from solar north) for the CME detected as cluster 4 in the left-hand panel of Figure 4. Right: same as for the left, but for cluster 5. The box-and-whisker keys indicate the median and quartiles of the velocity distribution, as well as the minimum and maximum velocity values (see text for details).

(HELCATS) project.[10] The manual catalog compiled as part of HELCATS work package 2 contains observational CME parameters (in particular, time of first observation and northern and southernmost PA extents) obtained through visual inspection of the *STEREO*/HI-1 images. This catalog is available online.[11] Indeed, the application of CACTus to the *STEREO*/HI-1 imagery documented here is also performed under the auspices of the HELCATS project. The HELCATS manual catalog is described, and preliminary analysis undertaken, by Harrison et al. (2016, in preparation). Moreover, we compare the CACTus POS speed estimates with those contained within the augmented (work package 3) version of the HELCATS manual CME catalog[12], which also includes estimates of 3D radial speeds calculated along, or at least near, the CPA of each CME. The 3D speeds in this augmented catalog are derived through the analysis of the time-elongation profiles of the CMEs manually tracked in time-elongation maps (J-maps) generated from combined HI-1 and HI-2 imagery, using three techniques: Fixed-ϕ Fitting (FPF; Rouillard et al. 2008), Self-Similar Expansion Fitting (SSEF) with an assumed angular half width of 30° (Davies et al. 2012) and Harmonic Mean Fitting (HMF; Lugaz et al. 2009). The FP, HM and SSE geometries that underlie these fitting techniques assume different values for the cross-sectional width of a CME. FP assumes a half width of 0°, i.e., CME is a point source, whereas HM assumes a half width of 90°, i.e., a circle tied to the Sun-center. The SSE geometry is more general, with a half width, which can be anywhere between 0° and 90°. 30° is chosen for use in the HELCATS manual catalogs to account for the fact that the average angular span of CME is around 60°. We reiterate here that the POS approximation used by CACTus is a special case of the FP approximation, assuming CME propagation to be at 90° from the Sun-spacecraft line.

3.1. Comparison of Number of Events Detected

During the 3 day period extending from the beginning of 2010 April 02 to the end of 2010 April 04, CACTus detected five CMEs (see, Figure 4 and Tables 1 and 2). Since we incorporated that three day period into a single CACTus analysis, we get only a single CME map (shown in Figure. 4), which contains all events detected during that period. The other 5 days were analyzed individually. Table 2 compares the number of CMEs detected by CACTus with the number detected manually as a part of the HELCATS project. We find that CACTus detected more events than were listed in the manual catalog during the 8 days analyzed (11 as opposed to 6). Over-estimation of the number of CMEs appears to be a common feature of CACTus (see Robbrecht et al. 2009). Having examined the CMEs automatically identified in the HI-1 data, we find that, in some cases, CACTus identifies multiple separate events, which, on closer investigation, are part of a single CME. *For example,* one event detected by CACTus on 2008 April 26 was found to actually be the flank of a separately detected CME. In that case, both the apex and the flank of the CME were bright, while the region between them was faint, hence their detection by CACTus as two separate events. The CME identified by CACTus on 2010 April 04 (Figure 5 (bottom panel)) is not listed in the HELCATS manual CME catalog. This is probably because this CME is narrower than the 20° PA width threshold imposed in the HELCATS manual cataloging. Similarly, the three events detected by CACTus on 2010 April 02 (CMEs 1, 2, and 3 in Figure 4), all are narrow unstructured features propagating radially outward (these are still identified as CMEs by CACTus according to the definition implemented therein) and are, therefore, not listed in the HELCATS manual catalog. In this work, we have set a

[10] http://www.helcats-fp7.eu
[11] http://www.helcats-fp7.eu/catalogues/wp2_cat.html
[12] http://www.helcats-fp7.eu/catalogues/wp3_cat.html

Table 1
Extract of the CACTus CME Catalog for 2010 April 02–04

No.	t_0	CPA	da	NoPA	SuPA	v	dv	minv	maxv
1	2010 Apr 02 05:29	116	12	110	122	459	73	316	498
2	2010 Apr 02 14:09	83	10	78	88	412	25	384	454
3	2010 Apr 02 18:09	104	16	96	112	319	65	229	397
4	2010 Apr 03 12:09	102	72	66	138	823	108	571	1041
5	2010 Apr 04 00:49	124	24	112	136	397	86	350	571

Note. No. represents the cluster number (CME number), t_0 is the time of appearance of the CME in the HI-1 FOV, CPA is the central position angle (PA) of the CME (degrees), da is its PA width (degrees), NoPA is its northernmost PA extent (degrees), SuPA is its southernmost PA extent (degrees), v is the median of the maximum value of POS velocity at each PA of the CME (km s^{-1}), dv is the (1σ) variation of the velocity over the entire PA extent of the CME, minv is lowest value of the maximum velocity and maxv is highest value of the maximum velocity.

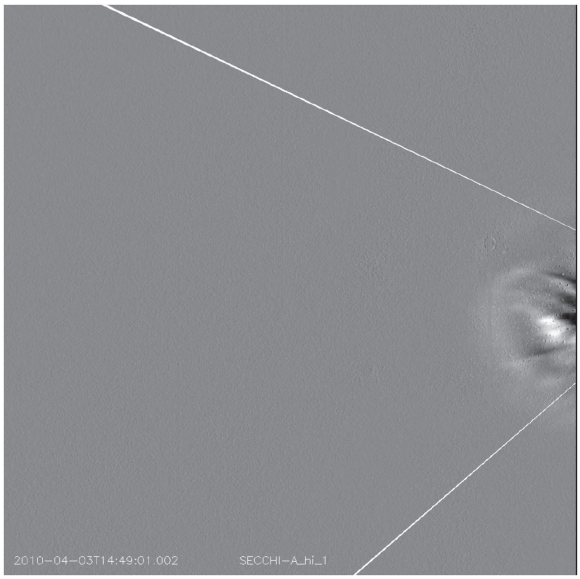

Figure 7. Movie corresponding to this animated figure shows the CME moving outward antisunward. The angular width of the CME is delimited by white lines.

(An animation of this figure is available.)

Figure 8. Movie corresponding to this animated figure shows a narrow CME moving outward. The angular width of the CME is delimited by white lines.

(An animation of this figure is available.)

Table 2
Comparison of the Number of Events

Date	Events Detected by CACTus	Events Detected in Manual Catalog
2010 Apr 04	1	0
2010 Apr 03	1	1
2010 Apr 02	3	0
2008 Dec 12	1	1
2008 Apr 26	2	1
2008 Feb 04	1	1
2007 Oct 22	1	1
2007 Apr 19	1	1

threshold of 10° in PA width for a feature detected by CACTus to qualify as a CME; application of CACTus to coronagraph images generally uses 5°.

In summary, CACTus detected all CMEs present in the manual catalog, although one of them was erroneously identified by CACTus as two separate events. Moreover, use of a 10° (rather than 20°) PA width threshold resulted, over these 8 days, in the identification by CACTus of four additional, narrow, CMEs. It could be argued that such narrow CMEs, which (presumably they are small blobs of plasma) usually follow a larger CME, ought not to be categorized as a separate CME, but as a part of the preceding CME. In manual cataloging, such association is at the discretion of the operators. CACTus applies a more objective definition; any feature (with brightness exceeding $W_{\rm thresh}$), which appears more than 200 minutes after the appearance of a CME, is considered as a separate CME. Although this is the objective application of what is still a subjective threshold, it means that CME detection will remain consistent over a long period of time, provided that there are no significant changes in instrument performance (see, Tappin et al. 2016).

3.2. Comparison of Angular Width

We also note that CACTus estimates consistently lower PA widths for the CMEs detected during the days under study than those presented in the HELCATS manual catalog (see, Table 3). We suggest that a possible reason for this is that CMEs expand as they propagate outward. Moreover, the flanks of a CME are often fainter than its leading edge. Consequently, the ridge corresponding to the flank of a CME in a time–height map at a given PA may not be sufficiently extended or bright to result in its detection by CACTus (i.e., its weight may not exceed

Table 3
Comparison of CACTus vs. Manual Detection

CACTus				Manual						
t_0	CPA	da	v	t_0	CPA	da	v_{SSE}	ϕ_{SSE}	v_{FP}	v_{HM}
2010 Apr 03 12:09	102	72	843	2010 Apr 03 12:09	102	105	927	77	889	962
2008 Dec 12 15:29	77	70	469	2008 Dec 12 15:29	77	95	426	69	419	431
2008 Apr 26 18:09	82	72	595	2008 Apr 26 18:49	85	110	649	95	620	679
2008 Feb 04 14:09	83	66	448	2008 Feb 04 14:09	77	115	511	53	506	515
2007 Oct 22 14:49	84	68	225	2007 Oct 22 14:49	77	85	691	174	499	748
2007 Apr 19 11:30	90	72	376	2007 Apr 19 13:30	90	100	392	61	389	393

Note. v_{SSE} is the speed in km s^{-1} and ϕ_{SSE} is the spacecraft-Sun-CME angle in degrees derived using SSEF (Davies et al. 2012). If ϕ_{SSE} is close to 90° then the CME originated from near the limb. If ϕ_{SSE} is close to 0° or 180° then CME is a front or backsided halo, respectively, as seen from the *STEREO-A* spacecraft. v_{FP} and v_{HM} are the speed in km s^{-1} derived using FPF and HMF, respectively.

W_{thresh}). Another possible reason for the discrepancy in manually and automatically deduced PA widths could be related to the fact that for fast CMEs in particular, the CME is often surrounded by a shock that tends to be fainter than the CME itself. We believe that it is more likely that a manual cataloger would be influenced by the presence of even faint shocks when estimating the CME width. Depending upon the brightness of the shock, it may (or may not) be detected by CACTus. It is debatable whether a shock or any structures adjacent to the CME, which are deflected due to the presence of a shock, should (or should not) be included when estimating the width of a CME. However, we believe that the CACTus results would only be affected by the presence of the brightest shocks (i.e., those with the longest tracks in a time–height map). It should be pointed out that the discrepancy in the PA width estimation is also seen when comparing the width of CMEs detected by CACTus in coronagraph data with those detected manually (Robbrecht et al. 2009). Thus, we believe that the estimation of the PA width of CMEs in HI-1 data is affected by: (a) possible non-radial motion of CMEs, (b) the tendency of CME flanks to be faint and hence their ridges in time–height maps to be of limited extent and (c) the presence of faint shocks around CME flanks. Potentially, the PA widths estimated by CACTus could be made more consistent with the values in the manual catalog by lowering the intensity threshold, but doing so would also increase the number of false detections.

3.3. Comparison of Time of Appearance in the HI-1 FOV

We find that the times of first appearance (t_0) of the CMEs in the HI FOV, as estimated by CACTus and in the HELCATS manual catalog, are in fairly good agreement (see Table 3), with the exception of the event on 2007 April 19. For that event, CACTus yields a time of entry into the HI-1 FOV, which is two hours earlier than the value quoted in the manual catalog. Having examined the images, we believe CACTus to be correct; the CME is already well inside the HI-1 FOV at the time of first appearance quoted in the HELCATS catalog. The reason for this discrepancy is not clear. Nevertheless, it confirms the rather subjective nature of manual cataloging. In general, it is thought that such time differences are due to (a) narrow blobs preceding a CME (Robbrecht et al. 2009) or (b) the inclusion of the shock ahead of a CME. This is discussed in greater detail in Section 4.

3.4. Velocity Comparison

We have compared the POS velocity from CACTus with the best-fit 3D velocity derived using FPF (v_{FP}), SSEF (v_{SSE}) with an assumed angular half width of 30°, and HMF (v_{HM}). The speeds are quoted in Table 3. ϕ_{SSE}, also listed in Table 3, refers to the best-fit angle between the Sun-observer line (in this case *STEREO-A*) and the propagation direction of the CME returned by the SSEF technique (ϕ_{FP} and ϕ_{HM} are also quoted in HELCATS catalog, and for the CMEs considered here are similar in value to ϕ_{SSE}). Since SSE assumes a more generalized geometry, which reduces to FP and HM geometries under certain limits, we have quoted only ϕ_{SSE} for the CMEs considered in Table 3. In particular, we compare the POS velocities from CACTus with those derived using the FPF technique, because the POS geometry is a special case of the FP geometry.

Note that ϕ_{SSE} is a reasonable proxy for the longitudinal separation between the CMEs and the spacecraft. We find that the POS speed from CACTus and v_{SSE} match reasonably well, except for the CME on 2007 October 22. We note that ϕ_{SSE} for this event is 174°, meaning it propagated more or less directly away from the spacecraft. Therefore, its projected speed in the POS would be a gross underestimate of its 3D radial speed. The POS approximation underestimates the 3D propagation velocity of any CME, which is propagating in a direction well away from the limb (e.g., a halo CME). For all other events, we find that ϕ_{SSE} is close to 90° (i.e., the CMEs are near-limb events) and so it is not surprising that the POS speed is more consistent with the 3D speed derived using the FPF, HMF, and SSEF techniques. We also find that, for these near-limb events, the velocities derived using FPF, HMF, and SSEF (which are based on three different underpinning geometries) are in fairly good agreement. Remember, for a limb event, the FP geometry is equivalent to the POS approximation. It is worth repeating here that, with the current implementation of CACTus we do not use any a priori information regarding a CME's propagation angle and we cannot estimate the propagation direction in the same way as FPF, SSEF, and HMF techniques as we cannot detect any curvature of the ridges in a time–height map. However, the POS approximation is a good approximation for near-limb events.

4. ONLINE AUTOMATED CATALOG OF SECCHI/HI-1 GENERATED BY CACTUS

In the sections above, we initially selected eight different days (a three day run, and five individual days) to demonstrate

Table 4
Comparison of Online CACTus Catalog with Manually Generated Catalog

CACTus Online				Manual							
t_0	CPA	da	v	t_0	CPA	da	v_{SSE}	ϕ_{SSE}	v_{FP}	v_{HM}	Type
2013 May 13 20:09	87	94	745	2013 May 13 20:09	87	135	738	81	705	769	good
2013 Feb 27 15:29	83	82	411	Extra event identified by CACTus							
2013 Feb 27 10:09	96	88	520	2013 Feb 27 10:09	95	130	597	35	577	612	fair
2012 Jul 02 21:29	115	42	502	2012 Jul 02 20:49	100	100	434	75	426	441	good
2012 Feb 06 02:49	67	46	414	2012 Feb 06 02:49	63	65	410	95	399	420	fair
2011 Jun 01 22:09	88	60	505	2011 Jun 01 22:49	92	85	774	139	667	931	good
2011 Feb 26 04:09	82	64	352	2011 Feb 26 05:29	72	85	389	48	383	392	fair
2010 Jun 03 11:29	88	64	307	2010 Jun 03 12:09	90	120	333	100	320	347	good
2010 Apr 03 12:09	102	72	823	2010 Apr 03 12:09	102	105	927	77	889	962	good
2009 Jul 31 12:49	82	36	334	2009 Jul 31 13:29	40	40	331	70	326	334	fair
2009 Jan 09 07:29	92	48	330	2009 Jan 09 08:49	88	65	522	26	480	570	good
2008 Dec 12 14:09	77	78	475	2008 Dec 12 15:29	77	95	426	69	419	431	good
2008 Apr 26 16:49	91	98	553	2008 Apr 26 18:49	85	110	649	95	620	679	good
2008 Feb 04 14:09	96	40	437	2008 Feb 04 14:09	77	115	511	53	506	515	good
2008 Feb 04 03:29	75	30	505	Not listed in catalog							
2007 Oct 22 19:29	74	24	224	Single event is detected as two different events by CACTus							
2007 Oct 22 14:49	92	20	218	2007 Oct 22 14:49	77	85	691	174	499	748	fair
2007 Apr 19 11:30	91	70	371	2007 Apr 19 13:30	90	100	392	61	389	393	good

the performance of the revised CACTus algorithm. Using this version of CACTus, we have also generated a full CME catalog, extending from 2007 to 2014, which is available at http://www.sidc.be/cactus/hi and http://www.helcats-fp7.eu/catalogues/wp2_cactus.html. The methodology employed to generate the full catalog (i.e., the thresholds in intensity, etc) is identical to what has been described above, except that the analysis is performed on intervals of one month duration, which overlap by one day in order to accommodate CMEs, which span the month boundary. A limit of one month is imposed for computational reasons. It is important to note that the length of time interval, which we incorporate into a single CACTus analysis run, will affect the detection and/or returned parameters of a CME, since the intensity threshold in accumulator space (W_{thresh}) depends on W_{mean} and W_{sd}, both of which depend on the duration of the analyzed interval. We choose 15 different "good" and "fair" quality CMEs from the HELCATS manual catalog, from different times between 2007 and 2013 and with a range of values of ϕ_{SSE} (see Table 4). Table 4 also includes the events listed in Table 3 (shown in boldface).

1. *Comparison with CMEs discussed in Section 3*: To demonstrate the effect of the length of the analyzed interval, we compare the results for the six CMEs listed in Table 3 with their entries in the online CACTus catalog. The dates and times of these CMEs are highlighted in boldface in Table 4. We note that there are, obviously, some differences that arise when an interval of one month is incorporated in a single analysis.
   (a) Of the six CMEs listed in Table 3, the times of appearance of four match well with their counterparts listed in Table 4 (in boldface). For the other two CMEs, there is a difference of two frames (corresponding to 80 minutes). As discussed above, this may be attributed to the different thresholds resulting from the analysis of different lengths of data.
   (b) Out of the six CMEs listed in Table 3, the PA widths of the CMEs detected on 2010 April 03 and 2007 April 19 match well with those listed in Table 4. For the CMEs detected on 2008 February 04 and 2007 October 22, the PA width quoted in Table 4 is less than the PA width quoted in Table 3. On closer inspection, we find that, when using a longer analysis interval, each of these CMEs was identified as two separate (narrower) CMEs. Reanalysis results in the CME on 2008 February 04 being identified as two separate CME, because the leading edge of the CME is poorly defined. We suggest that, in this case, the slight change in the intensity threshold due to the longer analysis interval results in the points in accumulator space not being joined up under the application of the morphological closing technique. Hence, the algorithm detects two narrower CMEs instead of a single wider one. Finally, for the CMEs detected on 2008 December 12 and 2008 April 26, the PA widths of the CMEs in Table 4 are greater than their counterparts in Table 3. This can also be explained in terms of there being a different threshold due to increased length of the analyzed data set, which results in the inclusion of additional material (possibly the CME flanks or adjacent structures). For these two CMEs, the PA widths quoted in the online CACTus catalog are actually much more consistent with those in the manual catalog.
   (c) Different lengths of analysis intervals will clearly also result in different CME velocities. CACTus computes the maximum velocity in each PA bin within a CME. The velocity listed in the online CACTus catalog is the median of all these maximum velocities. The median is likely to change if the PA width of the CME

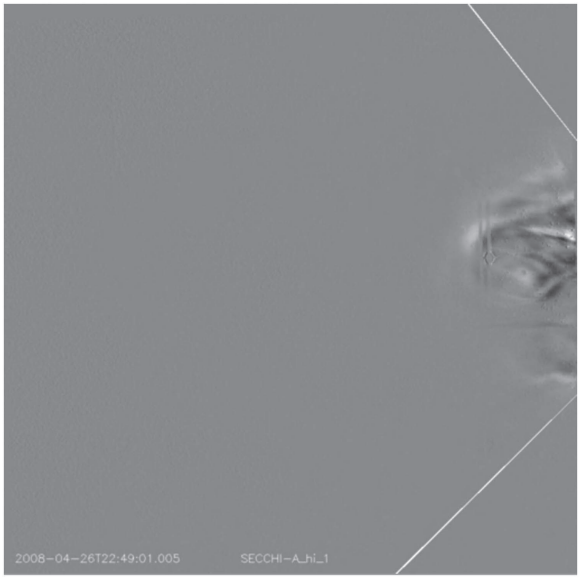

Figure 9. Movie corresponding to this animated figure shows the CME moving outward. Clearly seen pre-CME outflows, which affect the estimation of the time of appearance in HI-1 FOV.

(An animation of this figure is available.)

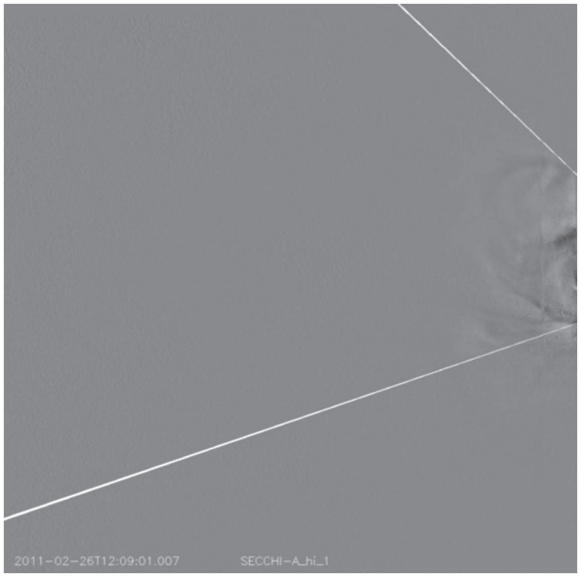

Figure 10. Movie corresponding to this animated figure shows a CME with faint leading edge moving outward. The faint and unstructured leading edge makes the estimation of time of appearance of a CME in HI-1 FOV difficult.

(An animation of this figure is available.)

changes. Moreover, a different threshold is likely to give rise to a different estimate of the maximum velocity at a given PA (see Section 2.4). It is likely that both of these effects would contribute to differences in the CME velocities if the length of the analysis interval is modified significantly. However, we find that the differences in the velocities of the common CMEs in Tables 3 and 4 are within one standard deviation (dv, quoted in Table 1).

It is difficult to ascertain whether it is better to analyze individual days (or short intervals) or longer periods of data (such as a month). However, it is important to employ a consistent methodology. Therefore, for in the generation of the online catalog, we incorporate one month of data into each CACTus run from 2007 to 2014.

2. *Comparison with manual catalog*: Here, we compare the online CACTus catalog with the manual catalog. We compare every event listed in both manual and automated catalogs on the selected days. A detailed comparison between the manual and the automated catalogs will be presented in future studies.
   (a) The times of CME appearance in the two catalogs differ at most by three frames (2 hr). In Section 3.3, we have compared the manual and automatically derived parameters of the CME, which entered the FOV of HI-1 on *STEREO-A* on 2007 April 19. The times at which the CME on 2008 April 26 was first detected by CACTus is two hours earlier than its manual catalog entry. We suggest that this discrepancy is due to the outflow of associated material prior to the CME itself (see animated Figure 9 available online). Sometimes the leading edge of a CME is faint or not well defined, for example the CME on 2011 February 26, which makes it difficult to estimate accurately its time of first appearance in the FOV (see animated Figure 10 available online). For some CMEs, the times of appearance match very well in both the automated and manual catalogs. Such events tend to have bright, clear leading edges (for example the CME on 2013 February 27) and/or are preceded by outflows, which are faint enough not to be detected by CACTus (see animated Figure 7 available online). In general, we suggest that CACTus tends to detect CMEs either at the same time or slightly earlier than is estimated by manual operators.
   (b) In general, CACTus tends to produce a lower estimate of the PA width than that included in the manual catalog for the same event. As noted previously, this can be due to the misdetection by CACTus of a single CME as two narrower events or can be due to the fact that CMEs flanks can be faint and can sometimes move non-radially.
   (c) Velocities estimated by CACTus tend to match well with v_{FP} in the HELCATS manual catalog for limb/near-limb CMEs (those with ϕ_{SSE} close to 90°).
   (d) For the days considered in Table 4, we find that CACTus detects a total of 18 events. The manual catalog includes only 15 events, meaning that CACTus detects 3 events, which are not listed in the manual catalog. One of these additional events (on 2013 February 27) is due to CACTus identifying, as a separate CME, pre-CME outflows. Another is due to the erroneous splitting of a single CME into two (2007 October 22). Finally, one narrow CME is also identified, which is not listed in the manual catalog as it is narrower than the PA width threshold of 20° imposed in the generation of that catalog.
3. *Corotating Interaction Regions (CIRs):* CIRs are the regions of high-density plasma, which form at the interface between the fast and the slow solar wind by compression (Gosling & Pizzo 1999). Rouillard et al. (2008) and Sheeley et al. (2008a, 2008b) first demonstrated white-light imaging of CIRs in the heliosphere

using *STEREO*/HI. However, the authors suggested that HI was actually observing pre-existing streamer blobs, which had become entrained at the stream interface. Compression of the blobs at the stream interface means that they can be clearly tracked out to 1 AU and beyond. Although it is thought that this entrainment happens beyond the HI-1 FOV (see Plotnikov et al. 2016), we have examined the CACTus output to see if it has detected any of the CIR-associated blobs presented in the literature. We find that CACTus detects some, but not all, of these features. For example, CACTus detects two of the six CIR-associated blobs, which were tracked by Rouillard et al. (2010) between 2007 September 09 and 12, but none of the six CIR-associated blobs tracked by those authors between 2007 September 17 and 20. Similarly, Tappin & Howard (2009) detected the tracks of several CIR-associated blobs between 2008 November 14 and 17, only one of which was detected by CACTus. The CIR-entrained blobs detected by CACTus have PA widths, which are less than 20°. Therefore, a PA threshold of 20° would likely exclude many such blobs. However, it may also remove some features, which are associated with CMEs. The detection of CIR-associated blobs using CACTus will be the subject of a future study.

5. SUMMARY AND DISCUSSION

We have applied, with success, the CACTus algorithm, to automatically detect CMEs in images from the HI-1 camera on board *STEREO-A*. We performed the analysis on 1 day background-subtracted level 2 data. We have experimented with different background-subtraction methodologies and find it makes little difference because, ultimately, we use polar transformed running difference images for detection using the Hough transform. In summary, the different background-subtraction techniques do not significantly affect the performance of CACTus, especially for bright CMEs.

We have also applied this algorithm to HI-1 images from *STEREO-B,* but with less success. The HI-1 camera on *STEREO-B* suffers from pointing anomalies due, it is now thought, to a small amount of mechanical instability in the attachment of the camera to the HI instrument structure (Brown et al. 2009; Davis et al. 2012; Tappin et al. 2016). Given the large gradients inherent in the F-corona, which has a much larger white-light signal than that of CMEs, such pointing jitter can result in the presence of a residual F-coronal signal in the *STEREO-B* HI-1 imagery. This residual signal detrimentally impacts the automatic detection, in particular, of CMEs by CACTus. Hence, more advanced techniques are required to detect CMEs reliably in HI-1 images from *STEREO-B*. Here, we primarily present the technical details of the detection algorithm, which we implemented within CACTus. Of the 8 days initially studied here, 3 days (2010 April 02 to 04) were analyzed as a single run and the remaining 5 days were analyzed individually. Incorporating these 5 individually analyzed days (see Table 3) into monthly runs (see Table 4) makes little difference to the CACTus results. Any differences are due to the different thresholds, which result from the incorporation of one month data. We also compare a few additional events taken at random from the online CACTus catalog (these encompass both "good" and "fair" CMEs as defined by the HELCATS manual catalogers). We compared the automatically and manually derived CME properties (see Table 4).

We find that, at least over the limited number of days presented in the current paper, CACTus detects more events in HI-1 images than are detected by the manual cataloging. This is due to several factors, which are inherent in the method (the selection of the intensity threshold value and kernel size for morphological closing). In this study, some of the events detected by CACTus are excluded by human operators due to different thresholds in the PA width extent. Manual-cataloging endeavors need to apply a relatively large PA width threshold, which is not necessary with an automated detection routine such as CACTus, simply due to the arduous nature of the manual-cataloging process. It is worth noting that even when we apply objective definitions, there are instrumental factors, which come into play even in automatic CME detection. For example, the apparent brightness of a CME in white-light imagery is dependent to some degree on its location relative to the Thomson sphere. Also, the application of a PA width threshold in order to identify a CME is not entirely consistent as PA width is a projection of a CME's true width onto the POS. Conversely, there are some benefits to a manual as opposed to an entirely algorithmic approach. For example, what is clearly a single CME in the HI imagery on 2007 October 22 (and in the manual catalog) is detected as multiple (two separate) events by CACTus due to the large variation in the brightness over the CME front. In fact, angularly extended faint CME fronts (as characterized in the manual catalog) commonly appear as several narrower CMEs in the CACTus catalog. This tendency for CACTus to overestimate the number of CMEs is also seen in the LASCO and COR2 implementations of CACTus, so it is an intrinsic problem of the method (Robbrecht et al. 2009).

We also note that CACTus sometimes computes the time of appearance of a CME to be earlier than it is in the manual catalog. This can be explained by the presence of pre-CME outflows or the lack of a well defined and bright leading edge. We feel that the differences in estimation of the time of first appearance of a CME do not necessarily mean that an automated algorithm is superior to the discretion of manual operators. Rather, it indicates the lack of a proper definition of the time of appearance of a CME.

We find that the PA width of a given CME estimated by CACTus tends to be lower than is listed in the manual catalog. This effect can be exacerbated by the presence of a faint shock around the CME, which is often taken into account by manual catalogers, but excluded most of the time by CACTus. This discrepancy in CME width was also present in the earlier version of CACTus (see Robbrecht et al. 2009). The difference in width measurement is also due to the lack of a proper definition of the width of a CME. Cremades & Bothmer (2004) reported a difference of a few degrees to more than a hundred degrees when comparing manual catalogs with their measurements. It is still subjective whether or not shocks should be included when estimating the width of a CME. CACTus implements a somewhat more objective criteria for this but, still, the shocks around the flanks of CMEs may or may not be included. The time of CME appearance from CACTus is, on average, earlier than that quoted in the manual catalog. A comprehensive comparison with the manual catalog is underway and will be presented in a future paper.

Currently, CACTus yields only the velocity projected onto the POS as seen from the observing spacecraft. This should be

borne in mind when using the speeds derived by CACTus for specific research purposes. Options for retrieving the 3D speed from CACTus may be sought in the future, for example through the inclusion of a more realistic estimate of the propagation direction (such as from GCS or the CAT tools). However, this may pose potential problems, in particular for its realtime implementation. It may also be that CACTus could be optimized to extract the curved ridges in time-elongation maps from heliospheric imaging observations, which extend out to large elongation; these tracks contain information regarding the 3D propagation direction within their curvature.

One distinct advantage of automated detection is that it is not affected by the vagaries of human subjectivity and, hence, the detection of CMEs will be more consistent. As noted above, one disadvantage of this is that it generally results in the detection of many more events than are listed in the manually compiled catalog. However, the number of detected events depends critically on the threshold value of brightness implemented in the program. Optimization of this value of this threshold for HI imagery will be evaluated in the future versions of CACTus for HI.

To summarize, in this work we have described the basic detection algorithm for detecting CMEs in heliospheric imager data using CACTus for the first time. This is of great importance for heliospheric physics. To the best of our knowledge, this is the first report on the successful fully automated detection of CMEs in data from the *STEREO* heliospheric imager HI-1.

The authors thank the referee for his/her valuable in-depth comments, which have helped us to improve the manuscript. V.P. acknowledges the support of the entire CACTus team at the Royal Observatory of Belgium. This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 606692 [HELCATS]. We acknowledge support from the Belgian Federal Science Policy Office (BELSPO) through the ESA-PRODEX program and from the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office (IAP P7/08 CHARM).

REFERENCES

Barnard, L., Scott, C., Owens, M., et al. 2014, SpWea, 12, 657
Barnard, L., Scott, C. J., Owens, M., et al. 2015, SpWea, 13, 709
Bisi, M. M., Harrison, R. A., Lugaz, N., van Driel-Gesztelyi, L., & Mandrini, C. H. 2013, SoPh, 285, 1
Boursier, Y., Lamy, P., Llebaria, A., Goudail, F., & Robelus, S. 2009, SoPh, 257, 125
Brown, D. S., Bewsher, D., & Eyles, C. J. 2009, SoPh, 254, 185
Brueckner, G. E., Howard, R. A., Koomen, M. J., et al. 1995, SoPh, 162, 357
Byrne, J. P. 2015, JSWSC, 5, A19
Byrne, J. P., Morgan, H., Habbal, S. R., & Gallagher, P. T. 2012, ApJ, 752, 145
Cremades, H., & Bothmer, V. 2004, A&A, 422, 307
Davies, J. A., Harrison, R. A., Perry, C. H., et al. 2012, ApJ, 750, 23
Davis, C. J., Davies, J. A., Lockwood, M., et al. 2009, GeoRL, 36, L08102
Davis, C. J., Davies, J. A., St Cyr, O. C., et al. 2012, MNRAS, 420, 1355
Eyles, C., Davis, C., Harrison, R., et al. 2007, Proc. SPIE, 6689, 668907
Eyles, C. J., Harrison, R. A., Davis, C. J., et al. 2009, SoPh, 254, 387
Eyles, C. J., Simnett, G. M., Cooke, M. P., et al. 2003, SoPh, 217, 319
Gonzalez, R. C., & Woods, R. E. 2002, Digital Image Processing (Upper Saddle River, NJ: Prentice Hall)
Gopalswamy, N., Lara, A., Lepping, R. P., et al. 2000, GeoRL, 27, 145
Gosling, J. T., & Pizzo, V. J. 1999, SSRv, 89, 21
Harrison, R. A., Davies, J. A., Rouillard, A. P., et al. 2009, SoPh, 256, 219
Harrison, R. A., Davis, C. J., Eyles, C. J., et al. 2008, SoPh, 247, 171
Howard, R. A., Moses, J. D., Vourlidas, A., et al. 2008, SSRv, 136, 67
Howard, T. A., & DeForest, C. E. 2012, ApJ, 752, 130
Howard, T. A., Webb, D. F., Tappin, S. J., Mizuno, D. R., & Johnston, J. C. 2006, JGRA, 111, A04105
Hundhausen, A. J., Sawyer, C. B., House, L., Illing, R. M. E., & Wagner, W. J. 1984, JGR, 89, 2639
Inhester, B. 2015, arXiv:1512.00651
Jahne, B. 1997, Practical Handbook on Image Processing for Scientific Applications (Boca Raton, FL: CRC Press)
Kahler, S. W., & Webb, D. F. 2007, JGRA, 112, A09103
Lugaz, N., Vourlidas, A., & Roussev, I. I. 2009, AnGeo, 27, 3479
Manoharan, P. K., & Mujiber Rahman, A. 2011, JASTP, 73, 671
Morgan, H., Byrne, J. P., & Habbal, S. R. 2012, ApJ, 752, 144
Möstl, C., Amla, K., Hall, J. R., et al. 2014, ApJ, 787, 119
Olmedo, O., Zhang, J., Wechsler, H., Poland, A., & Borne, K. 2008, SoPh, 248, 485
Plotnikov, I., Rouillard, A. P., Davies, J. A., et al. 2016, SoPh, 291, 1853
Riley, P., Schatzman, C., Cane, H. V., Richardson, I. G., & Gopalswamy, N. 2006, ApJ, 647, 648
Robbrecht, E., & Berghmans, D. 2004, A&A, 425, 1097
Robbrecht, E., Berghmans, D., & Van der Linden, R. A. M. 2009, ApJ, 691, 1222
Rouillard, A. P., Davies, J. A., Forsyth, R. J., et al. 2008, GeoRL, 35, L10110
Rouillard, A. P., Davies, J. A., Lavraud, B., et al. 2010, JGRA, 115, A04103
Savani, N. P., Rouillard, A. P., Davies, J. A., et al. 2009, AnGeo, 27, 4349
Schwenn, R. 1996, Ap&SS, 243, 187
Sheeley, N. R., Jr., Herbst, A. D., Palatchi, C. A., et al. 2008a, ApJ, 675, 853
Sheeley, N. R., Jr., Herbst, A. D., Palatchi, C. A., et al. 2008b, ApJL, 674, L109
Socker, D. G., Howard, R. A., Korendyke, C. M., Simnett, G. M., & Webb, D. F. 2000, Proc. SPIE, 4139, 284
Tappin, S. J., Eyles, C. J., & Davies, J. A. 2016, SoPh, Submitted
Tappin, S. J., & Howard, T. A. 2009, ApJ, 702, 862
Tappin, S. J., Howard, T. A., Hampson, M. M., Thompson, R. N., & Burns, C. E. 2012, JGRA, 117, 5103
Thompson, W. T. 2006, A&A, 449, 791
Vourlidas, A., & Howard, R. A. 2006, ApJ, 642, 1216
Wang, Y.-M., & Colaninno, R. 2014, ApJL, 784, L27
Webb, D. F., & Howard, T. A. 2012, LRSP, 9, 3
Yashiro, S., Gopalswamy, N., Michalek, G., et al. 2004, JGRA, 109, A07105
Zurbuchen, T. H., & Richardson, I. G. 2006, SSRv, 123, 31