Validation of CME Detection Software (CACTus) by Means of Simulated Data, and Analysis of Projection Effects on CME Velocity Measurements

K. Bonte · C. Jacobs · E. Robbrecht · A. De Groof · D. Berghmans · S. Poedts

Received: 10 May 2010 / Accepted: 26 February 2011 / Published online: 24 March 2011 © Springer Science+Business Media B.V. 2011

Abstract In the context of space weather forecasting, an automated detection of coronal mass ejections (CMEs) becomes more and more important for efficiently handling a large data flow which is expected from recently-launched and future solar missions. In this paper we validate the detection software package "CACTus" by applying the program to synthetic data from our 3D time-dependent CME simulations instead of observational data. The main strength of this study is that we know in advance what should be detected. We describe the sensitivities and strengths of automated detection, more specific for the CACTus program, resulting in a better understanding of CME detection on one hand and the calibration of the CACTus software on the other hand, suggesting possible improvements of the package. In addition, the simulation is an ideal tool to investigate projection effects on CME velocity measurements.

K. Bonte $(\boxtimes) \cdot C$. Jacobs $\cdot S$. Poedts

C. Jacobs e-mail: carla.jacobs@wis.kuleuven.be

S. Poedts e-mail: stefaan.poedts@wis.kuleuven.be

K. Bonte · C. Jacobs · S. Poedts Leuven Mathematical Modeling and Computational Science Centre, K.U.Leuven, Leuven, Belgium

E. Robbrecht · D. Berghmans Royal Observatory of Belgium, Ringlaan 3, 1180 Brussel, Belgium

E. Robbrecht e-mail: eva.robbrecht@oma.be

D. Berghmans e-mail: david.berghmans@oma.be

A. De Groof European Space Agency, c/o Royal Observatory of Belgium, Ringlaan 3, 1180 Brussel, Belgium e-mail: anik.degroof@esa.int

Centre for Plasma Astrophysics, K.U.Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium e-mail: katrien.bonte@wis.kuleuven.be

Keywords Corona, models, observations · Coronal mass ejections, initiation and propagation · Instrumentation and data management · Magnetohydrodynamics

1. Introduction

Coronal mass ejections (CMEs) are giant plasma bubbles confined by magnetic field lines that are ejected from the Sun and travel through the heliosphere with an average or typical velocity of 450 km s⁻¹, but this velocity can vary between 100 km s⁻¹ and 3000 km s⁻¹. Although the solar corona has been observed for thousands of years (during solar eclipses), the existence of CMEs was discovered only in the space age, viz. in the early 1970s. The first observations of these dynamic events were done by a coronagraph aboard the seventh Orbiting Solar Observatory (OSO-7) mission from 1971 to 1974 (Tousey, 1973). Since then thousands of CMEs have been observed by several space-born and ground-based coronagraphs covering about two full solar activity cycles. Since December 1995, the time when the SOHO spacecraft became operational (Domingo, Fleck, and Poland, 1995), the Sun was observed almost continuously and the number of CME events reported in event catalogues raised drastically. The characteristics of CMEs and statistical analysis of CME properties have been published in literature by many authors (see, e.g., Howard et al., 1985; Hundhausen, Burkepile, and St. Cyr, 1994; St. Cyr et al., 2000; Yashiro et al., 2004; Webb et al., 2006; Robbrecht, Berghmans, and Van der Linden, 2009). Detection was first done manually, but since quite a few years also automated detection became possible thanks to the development of CME detection tools, among which is the CACTus software, developed at the Royal Observatory of Belgium (ROB) (Robbrecht and Berghmans, 2004).

The more observations confirmed the complexity of coronal mass ejections and their possible geo-effectiveness, the larger became the need for mathematically modeling these events in order to better understand and predict the effects of solar activity (on Earth). There are different kinds of CME models. On one hand there are magnetohydrodynamic (MHD) simulations which can provide in-depth physical information and a deeper insight in the global structure and propagation properties of these impressive solar events. Several MHD models have been developed and have helped in better interpreting observations. See for example Jacobs et al. (2009), Cohen et al. (2009), and Lugaz et al. (2009) for some recent 3D simulations studying CME initiation and propagation through the heliosphere. On the other hand also heuristic simulations are available, based on different geometrical shapes (Thernisien, Howard, and Vourlidas, 2006; Ciaravella, Raymond, and Kahler, 2006; Liewer et al., 2006; Boursier, Lamy, and Llebaria, 2009; Mierla et al., 2009). These simulations are limited by their physical content (density, pressure, background solar wind, internal CME structure, ...) but, due to their simplicity, they can be very efficiently produced and easily handled. They are ideal to study projection effects but do not suffice to simulate the intensities and densities observed in real white-light images. In this paper we aim at doing a complementary study in analyzing only one MHD-simulated CME event (with one particular geometrical shape) but containing the physics needed to construct a realistic white-light CME.

There still is a gap between two key elements in space weather predictions, namely between observational data of the solar corona and the solar wind on one hand, and numerical simulations of magnetic plasma clouds erupted from the solar corona on the other hand. There are several intrinsic and important differences between the numerical CME simulations currently being developed all over the world, and the latest coronagraph observations of CMEs. Coronagraph images show real coronal mass ejections, in a wide range of different shapes, velocities, and effects on Earth. But they are in general two-dimensional, limited through projection effects, and they only show the CME from one viewpoint (although some 3D information can now be extracted from simultaneous observations by the STEREO instruments (Kaiser *et al.*, 2008; Aschwanden *et al.*, 2009)), and suffer from a whole range of instrumental limitations like noise, limited temporal and spatial resolution, and even from the space weather effects they want to study. In addition, no physical characteristics can be directly read from the observational images. For numerically simulated CMEs, all physical characteristics (density, magnetic field, pressure, velocity, ...) are known in three spatial dimensions and at certain moments in time (the temporal resolution can be chosen), up to a predefined distance from the Sun. However, the models that are used for the initiation and propagation of these plasma clouds are idealized, in order to simplify the extremely complex structure of the corona and the solar wind conditions.

The aim of this work is to reduce the gap between these two completely different views of CME events. We believe this can be done from two different perspectives: first, as we discuss in this article, by the scientific visualization and interpretation of 3D numerical CME simulations mimicking the way the current solar coronagraphs observe such a plasma cloud, and using the automated way these observations are interpreted. Secondly, observations of the solar corona could improve the quality of numerical simulations by using observational data as input to obtain very realistic boundary conditions in the (existing) numerical codes.

The detection software CACTus is described in more detail in the next section. To validate the detection software, we technically preprocessed one specific CME-simulation dataset, so that it could be interpreted as coronagraph data by CACTus (Section 3). The 3D MHD simulation yielding the dataset is discussed in some detail in the Appendix. The results of our research are summarized in Section 4. The simulation is also an ideal tool to investigate projection effects on CME velocity measurements, as discussed in Section 5. We conclude this paper in Section 6 where we pose some ideas for further improvement and exploitation of the detection software and formulate our conclusions.

2. CACTus

CACTus stands for "Computer Aided CME Tracking" software. The package was first reported in Berghmans (2002) and is more extensively described in Robbrecht and Berghmans (2004). Its output is available online at http://sidc.be/cactus. CACTus is designed to detect CMEs in coronagraph images. Introducing automated CME detection in addition to "manual" detection, is important for efficiently handling a large data flow which is expected from recently-launched and future solar missions. The data flow can be processed faster and more continuously, which is important for the space weather community, *e.g.* for timely detection of halo CMEs. Moreover, automatically detecting CMEs is far more objective, compared with manual detection. For example, the CACTus CME rate shows a higher correlation with the solar cycle than the 'manual' CME rate (Robbrecht, Berghmans, and Van der Linden, 2009). Here, we provide a short description of how the detection software works.

The input to the CACTus software is a time sequence of coronagraph images. So far the detection software is running on observational data from the LASCO C2 and C3 coronagraphs (SOHO), and from the SECCHI COR2 coronagraph (STEREO). When working with LASCO coronagraph data, the software merges C2+C3 images to cover the combined field-of-view. The field-of-view of an instrument is the radial restriction of the view. In this case of coronagraph images, the field-of-view is restricted by the minimal r_{min} and the maximal



Figure 1 Transformation from the original observed coronagraph image in Cartesian coordinates I(x, y) to the polar-coordinate image $I(\alpha, r)$. The package merges C2 and C3 data.

 r_{max} distance from the Sun covered. The cadence of a dataset is the time between two images taken by an instrument. The field-of-view and cadence for the mentioned instruments are as follows:

- C2: $1.5 6 R_{\odot}$, 20 min
- C3: $4 30 R_{\odot}$, 30 min
- COR2: $2.5 15 R_{\odot}$, 15 min

These observational data are provided in the Flexible Image Transport System (FITS) file format (Ponz, Thompson, and Munoz, 1994), where the pixels contain actual intensity values. The time sequences are further referred to as datacubes of intensities: I(x, y, t).

Similar to other automated feature recognition tools, the CACTus algorithm consists of three steps:

- *i*) the preprocessing of the data;
- ii) the feature extraction; and
- iii) the output generation.

During the preprocessing, the data are cleaned from cosmic rays, instrumental stray light, and F-corona backgrounds. In addition, the coronagraph images are resized and transformed to a polar-coordinate system (α, r) , where α denotes the poloidal angle around the Sun and r corresponds to the radial distance from the limb. See Figure 1. This results in datacubes $I(\alpha, r, t)$, with r from r_{min} to r_{max} and t up to the duration of the dataset. To enhance the CME signal and reduce the background intensity, the software works with running difference images.

After preprocessing the original data, time–height (t, r) diagrams are created for each angle α : $I_{\alpha_i}(t, r)$, see Figure 2. The unique approach of CACTus is that it detects CMEs in these time–height diagrams. In such type of diagram, each solar event appears as a J-curve or an almost straight line of higher intensity, the so-called ridge, as clearly illustrated in Figure 2. The slope of the ridge indicates the speed of the event. Since the acceleration of a CME occurs close to the solar limb, these curves can very well be approximated with straight lines in the field-of-view up to $30 R_{\odot}$, *i.e.* assuming constant CME propagation. These bright lines are searched for by using the Hough transform (Ballester, 1994; Jahne, 1997). The latter method makes use of "template matching" to recognize objects within a



Figure 2 Extracting the time-height (t, r) diagram per angle α_i from the sequence of (α, r) images. The horizontal range corresponds to the duration of the time sequence. The vertical range corresponds to the combined C2+C3 field-of-view.

certain class of shapes by a voting procedure. Given an image $I_{\alpha_i}(t, r)$, the Hough transform $I_{\alpha_i}(a, b)$ is constructed by evaluating the integral

$$I_{\alpha_i}(a, b) = \int I(t, at + b) dt,$$

$$a = \Delta r / \Delta t,$$

$$b = -(\Delta r / \Delta t) * t_0 + r_{\min}$$

$$\Delta r = r_{\max} - r_{\min},$$

where Δt is the time period corresponding to Δr .

If a bright line $r = a_j t + b_j$ is present in the image $I_{\alpha_i}(t, r)$, then the Hough transformed image $I_{\alpha_i}(a, b)$ will have a local maximum at $[a = a_j, b = b_j]$. Detecting ridges in the images $I_{\alpha_i}(t, r)$ thus comes down to thresholding peaks in the transformed images $I_{\alpha_i}(a, b)$. In the case of CACTus, straight lines are characterized by two parameters t_0 and Δt in the (t, r) diagrams, corresponding to the coordinate of the intersection point with the time-axis (onset-time) and the time period corresponding to a distance $r_{\text{max}} - r_{\text{min}}$ in the radial direction, respectively. The modified Hough transform of a line is a point in the "Hough Space" with intensity value equal to the so-called ridge intensity. This is the Hough transformed value calculated for this ridge; see Figure 3.

CACTus repeats the Hough transform detection process on the (t, r) diagram for every angle. For each angle, the Hough procedure iterates over all possible values of t_0 and Δt , relating all possible ridges in the (t, r) diagram with corresponding points in the Hough space. Points of interest in the Hough space, *i.e.* points with peak values of integrated intensity, are grouped together using clustering and morphological closing operations to mark out different events. When an iteration threshold is reached, the points of interest are calculated back to the corresponding straight line(s) in the (t, r) diagram (visualized in Figure 4), from



Figure 3 Explanation of one Hough iteration. This figure shows that a straight line in the (t, r) diagram after Hough transformation corresponds to one point in the Hough space. The value corresponding to this point in the Hough space is the line integral of intensities.



Figure 4 Applying the Hough transform method to detect bright ridges in the $I_{\alpha_i}(t, r)$ image (top left). Below we visualize the corresponding Hough space, *i.e.* the Hough transformed image $I_{\alpha_i}(a, b)$. From there we threshold peak values of ridge intensity in the transformed image, an iterating process. At top right we visualize the inversion back to the detected ridges in the time–height (t, r) diagram after *n* iterations.

which the velocity is calculated. Since CACTus assumes constant velocity, no acceleration is measured. However, because it detects CMEs independently in each radial direction, a speed profile over the width of the CME is obtained. As such also the angular width of a CME can be measured.

After detection, an event is categorized as a genuine CME, or as a "flow". A flow is defined as a marginal, unimportant, or false detection. This categorization is based on the following restricting criteria: the velocity should be between 100 km s^{-1} and 2000 km s^{-1} , the apparent angular width has to be above 7°, the standard deviation in velocity has to be larger than zero and the ridge intensity should be above a certain threshold. Exactly this characterization is one of the weaker points of automated CME detection. In subtle cases as (partial) halo CMEs, a human observer's interpretation can be quite important for correctly characterizing an event.

A typical output of CACTus shows the detected events in an (α, t) diagram, including per event characteristics like the onset time, the principle angle (the projection of the principle direction of an event on the field-of-view), the angular width, and the velocity estimation.

3. Simulating Coronagraph Images

The 3D MHD simulation of the solar corona and the propagation of a CME used here to validate the CACTus software is described in Poedts *et al.* (2009). The simulation domain covers the lower solar corona and extends well beyond $30 R_{\odot}$. The simulation provides information about the plasma density, the momentum, the energy density, and the magnetic field vector at each grid point. A total of 14 artificial white-light images were retrieved from the simulation, with a time cadence of 15 minutes. The simulation was originally configured for investigating the propagation of CMEs throughout the inner heliosphere, applying a rather simple initiation of the CME. Therefore, we chose not to take into account the field-of-view below $4 R_{\odot}$, and to focus on the LASCO C3 field-of-view which suffices for validating the automatic detection package. For convenience, a detailed summary of the simulation setup and the results of the simulation is given in the Appendix.

In our study we use the synthetic data as input for CACTus instead of 2D white-light images of the Sun retrieved by coronagraphs. Therefore, white-light intensities are reconstructed from the numerical simulation output for each snapshot in time. At first instance, the observer is placed in the Sun–Earth Lagrange 1 point. In the spherical coordinate system (see Figure 5) used in the simulation, this corresponds to the observer being positioned at $(r, \theta, \varphi) = (212.7 R_{\odot}, 90^{\circ}, 90^{\circ})$, along the positive y-axis.

Important here is that we have to take into account that the light from the solar disk undergoes Thomson scattering due to the free electrons in the corona. To obtain the total line-of-sight brightness, the radiation scattered by all the electrons along the line-of-sight has to be integrated. The total intensity (K_{tot}) received by the observer is the sum of the intensity of the light scattered toward the Earth vibrating tangentially to the solar surface (K_t) and of the light vibrating radially (K_r). These formulas are well known (Minnaert, 1930) and are written down below, using the same notation as Wang *et al.* (1997):

$$\begin{split} K_{\text{tot}} &= \frac{\pi}{2} \sigma_{\text{T}} I_0 \int_{-\infty}^{+\infty} n_e \left\{ 2 \left[(1-u)C + uD \right] - \sin^2 \chi \left[(1-u)A + uB \right] \right\} \mathrm{d}s, \\ A &= \sin^2 \gamma \cos \gamma, \\ B &= -\frac{1}{8} \left[1 - 3\sin^2 \gamma - \frac{\cos^2 \gamma}{\sin \gamma} \times \left(1 + 3\sin^2 \gamma \right) \ln \left(\frac{1 + \sin \gamma}{\cos \gamma} \right) \right], \\ C &= \frac{4}{3} - \cos \gamma - \frac{1}{3}\cos^3 \gamma, \\ D &= \frac{1}{8} \left[5 + \sin^2 \gamma - \frac{\cos^2 \gamma}{\sin \gamma} \times \left(5 - \sin^2 \gamma \right) \ln \left(\frac{1 + \sin \gamma}{\cos \gamma} \right) \right], \end{split}$$

where σ_T is the Thomson scattering cross section, I_0 is the intensity of the solar radiation at the disk center, n_e is the electron density, u is the limb darkening coefficient, χ is the angle between a radial vector through the scattering point P and the line-of-sight, γ is the angle between a radial vector through P and a tangent from P to the solar surface, and ds is a path element along the line-of-sight. A value of u = 0.6 was chosen for the limb darkening coefficient.



Figure 5 Left: an observer close to the Earth observing the Sun over the line-of-sight (green line); the red arrow represents the CME with principal angle in the negative *X*-direction. Right: a sketch of the field-of-view (FOV), as seen from the Earth. In this field the principal direction of the CME corresponds to $\alpha = 270^{\circ}$.



Figure 6 Visualizations of running differences of the white light data of the simulated CME at three different times. Settings are chosen to simulate LASCO C3 coronagraph images, namely respecting a configuration where the field-of-view is limited to $30 R_{\odot}$ and where the light coming from the inner $4 R_{\odot}$ is blocked.

In Figure 6 we visualize three examples of the artificial white-light images of the simulated CME, achieved in the way as described above.

4. Analysis of the Simulated Data

Our approach is to consider the numerical simulation as a new instrument/detector, for which we install a function allowing the package to load the new type of data. The next steps are firstly to interpret the detections, and secondly to improve the software if necessary, knowing in advance that the simulated CME has the following properties:

- FOV = $4 30 R_{\odot}$
- Cadence = 15 min
- · The simulated dataset contains 14 images
- Principal direction $\alpha = 270^{\circ}$
- $\langle v_{270^\circ} \rangle = 1594 \text{ km s}^{-1}$

We constructed white-light coronagraph-like images from the simulated dataset, but we have to be aware that they show a different signal-to-noise level compared to real coronagraph images. In Figure 7 we show the signal after all the processing steps for an average CME, while Figure 8 shows the same for the CME resulting from the simulation. The signal is the running difference signal, scaled relative to the coronal background and the image cadence. As can be seen in the histograms on the left, the typical signal is on the order of a few percent of the background corona. CACTus truncates the signal at 5% to limit the influence of individual bright sources (planets, comets, etc.). On the right we show both a logarithmically scaled polar image and an image of all the pixels above 2%. In both the observed and the simulated cases, this amounts to about 1200 pixels. Nevertheless there are important differences between the simulated and the observed datasets. The observed dataset is dominated by noise, showing a power-law behavior in the histogram. In the observational data, debris or other objects may appear as bright strikes in the images. Moreover, the coronagraph itself is subject to space weather conditions: solar energetic particle (SEP) events can give rise to "snow" in the images. Still, after the preprocessing of the data, the observational datasets contain actual CME events buried in this noise. The simulated dataset is essentially noise free, except for some numerical fluctuations.

4.1. Generalizing Detection Thresholds

At first, the ideal, simulated CME (without instrumental noise) was not detected at all. The package as it existed was efficient in detecting, but some thresholds in the detection procedure of the CACTus code were hard-coded in the way that is only applicable to the observations currently used. They are, however, not suitable for images with a completely different signal-to-noise level. Therefore some technical modifications were applied to the code and





Figure 7 An example of an observed CME where the signal is truncated at 5%. Left: histogram of this dataset. Right: a pair of images scaled logarithmically (left) and linearly but truncated at the intensity level of 2% (right).



Figure 8 The same as Figure 7, but from the simulated CME.



Figure 9 Visualization of intensities in the angle-distance (α, r) diagrams at one particular time step where our simulated CME is expected to be detected. The horizontal range covers all angles α , from 0° to 360°, and the vertical range corresponds to the C3 field-of-view: $4 - 30 R_{\odot}$. Intensities that are detected as potential part of an event are marked by a random color per group of intensities. Left: applying the original parameters (optimized only for current observations) results in the wrong detection (detection of only the weakest intensity). Right: working with generalized detection thresholds improves the detection.

intensity thresholds were made relatively dependent on the overall background intensity instead of hard-coded to a fixed value. The result of these modifications is shown in Figure 9, which compares how the detection of the relevant intensities is made at one time step for all angles in an (α, r) diagram. The CME is clearly visible in the (α, r) image. However, the original detection missed most of the CME. Working with more general thresholds in the detection procedures significantly improved the result.

4.2. Limited Datasets: Consequences for Velocity Measurement

Secondly it became clear that the Hough transform, taking care of the calculation of velocities, results in less accurate measurements in cases where an event occurs before or in the very beginning of a dataset, as is the case for our simulated event. Working with our simulated CME, an average velocity of 2000 km s⁻¹ instead of 1594 km s⁻¹ was measured. The reason is that the parameter t_0 is restricted by the start of the dataset, which results in too steep a slope corresponding to the peak value of ridge intensity, and as such results in too high a measured speed. See Figure 10.

To prevent erroneous velocity calculations, it is sufficient to add 10 "neutral" images, without event, before the 14 images containing the simulated CME, embedding our simulated dataset in the environment of the quiet Sun. This embedding takes care that the parameter t_0 is not restricted by the limitation of the dataset. The improved result is shown in Figure 11. Another way to improve the Hough transform on this point could be to allow the parameter t_0 to become negative, but the solution of embedding our simulated data offers a better result also for other purposes (see the next section). When including the neutral images, an average velocity of about 1500 km s⁻¹ was measured for a principal angle of 270°, consistent with the velocity calculated in the Appendix. Adding even more neutral images did not influence the velocity measurements anymore.

4.3. Categorization of Detected Events

Apart from detecting, the CACTus software also distinguishes CMEs from flows in the detected events. As mentioned at the end of Section 2, this distinction is based on different restricting criteria: for an event to be classified as a genuine CME, it should have a realistic velocity, have a wide enough angular width, the velocity standard deviation has to be larger than zero, and the ridge intensity should be above a certain threshold. The first three of these four criteria are fulfilled. However, the CME event in the simulated dataset of 14 images was categorized as a flow.



Figure 10 Visualization of intensities of our simulated event in the time–height (t, r) diagram for a certain angle α_i . As mentioned in Section 2, the Hough transform searches for the brightest ridge in such a diagram. Here the black line, visualizing a detected ridge, indicates that the restriction of parameter t_0 results in too steep a slope. (Time is expressed by the number of images, and radius is expressed by the number of radial grid points, corresponding to the field-of-view of $4 - 30 R_{\odot}$.)



Figure 11 Angle-velocity (α , v) diagrams, giving the velocity associated with the highest ridge intensity for each angle. Left: velocity distribution with the event in the first several images of the dataset. Right: the result when adding neutral images before the event, *i.e.* when embedding our simulated dataset in the quiet Sun environment. The vertical box in the velocity boxplot contains the middle 50% of the measured speeds. The horizontal line in the box is the median speed (given in the CACTus output list). The whiskers at both ends indicate, respectively, the minimal and maximal detected speeds within 1.5 times the box length from the edge of the box. For a broader description of the boxplot, see Appendix A of Robbrecht and Berghmans (2004).

This distinction takes place after an event is detected and its location is identified. At this stage, the program concentrates on the subset of the dataset where the event occurs. It zooms into the detected event and compares the increased intensity with the "background" intensity of the dataset (which is calculated based on running average intensity). When the intensity of the event reaches a level higher than one standard deviation above the background intensity, the event is classified as a genuine CME. Adding even more neutral images than previously to the dataset was necessary to calculate this background intensity accurately. By adding 20 neutral images, the overall intensity of the simulated dataset is sufficiently balanced, resulting in a correct classification of the simulated CME.

Figure 12 Time-intensity diagrams. Top: diagram of a CME detected in observations. The solid line represents the background intensity, $\langle bg \rangle$, of the dataset. The lowest dashed line is the CME classification threshold value (background intensity + standard deviation). The darkest solid line represents the intensities of the detected event, marked as "CME-int". The asterisks point out the intensities of the detected event that are considered to be part of a genuine CME. Middle: diagram of the "isolated" simulated CME which is not recognized as a CME. Bottom: improved classification of the simulated CME. These diagrams illustrate the sensitivity of CACTus to the type of input and the consequences on event categorization.



The time-intensity diagrams in Figure 12 illustrate the sensitivity of the calculation of background intensity on the type of input. The intensity values shown in such diagrams correspond to the running difference image intensities $I(\alpha, r, t)$ averaged over the angular width of the concerning event and integrated over the radial distance.

This example has taught us that CACTus potentially misclassifies genuine CMEs as flows if the background intensity is not calculated accurately. This can possibly happen when the dataset is truncated (the CME is too much isolated in time) or if the background calculation is disturbed by nearby events. We expect that during solar maximum the categorization of events could be biased toward smaller events (since successive events might occur close in time).

4.4. Data Processing: Consequences for the Determination of the Principal Angle

The simulated CME is originally symmetric in the α -direction. Thanks to this *a priori* knowledge, we realize in this study that the CACTus processing can create small asymmetries when resizing the images, when extracting the ridges or when clustering/grouping the ridges in events (for more details, see Robbrecht and Berghmans (2004)).

In Figure 13 (α , *t*) diagrams are shown at different stages during the detection procedure. The processing clearly had an influence on the symmetry in the data array. The second, small



Figure 13 Angle-time (α , *t*) diagrams of the simulated CME at four different stages during the data processing. From panel A to B to C the software preprocesses the data and filters important intensities. In panel D, CACTus has grouped the values of detected intensities using clustering and morphological closing operations to mark out different events. Separation of events is visualized by assigning different (random) colors.

event detected is no longer concentrated around $\alpha = 270^\circ$, as it was so originally. This small event is a plasma "blob" that typically follows a CME in the same principal direction.

The described phenomenon has most significant influence on small events ("blobs" or "flows") only. However, it is good to understand that for this reason the principal angle measured for small events has a larger error margin compared to this parameter measured for large events (CMEs).

5. Projection Effects

In all results presented above, the observer was placed in the ideal position such that no projection effects occurred: the CME was seen by the observer as a limb CME with source region at W90 ($\alpha = 270^{\circ}$). Since we work with a simulated CME where we can choose the position of the observer, this study is an ideal occasion to focus on the projection effects.

For space weather prediction it is important to be able to estimate the arrival time of a CME on Earth. It is well known that CME velocity measurements are strongly biased by projection effects and several efforts have been done on improving the estimations of the true speeds of CMEs (see, *e.g.*, Vršnak *et al.*, 2007; Dal Lago *et al.*, 2004; Michałek *et al.*, 2004).

For creating the projection effect in our CME simulation, we kept the direction (θ_{CME} , φ_{CME}) of the CME constant, but changed the position of the observer in the simulation setup (*cf.* Figure 5). In the first instance we moved the observer toward the principal direction of the CME, but we keep the observer still positioned in the equatorial plane of the Sun:

$$\theta_{\rm obs} = 90^{\circ},$$

 $90^{\circ} \le \varphi_{\rm obs} \le 180^{\circ}$

where θ_{obs} and φ_{obs} are the heliographic co-latitude and longitude of the observer as illustrated in Figure 5. Note that with the observer located in the equatorial plane, in which as



Figure 14 Angle–velocity (α, v) diagrams for different positions of the observer. The observers positions are, from left to right, $\varphi_{obs} = 90^{\circ}$, 130°, and 180°, respectively ($\theta_{obs} = 90^{\circ}$ for all of them). The velocity profiles clearly show the projection effect. The velocity boxplot is drawn as explained in Figure 11.

well the center of the CME propagates, the detected principal angle α of the CME will always be 90° or 270° except in the limit of $\varphi_{obs} = 180^{\circ}$. In the latter case the CME travels straight along the observer's line-of-sight, resulting in a halo CME.

Figure 14 clearly shows that, although we always consider the same CME, moving the observer's position by increasing φ_{obs} indeed results in a reduction of the measured velocity, since the velocity vector is projected on the plane of the sky. The closer the observer is moved toward the direction of the CME, the more the velocity values flatten out for all angles. Placing the observer exactly in the principal direction of the CME ($\varphi_{obs} = 180^\circ$) results in similar velocity values measured in all directions: the package has a good interpretation of a halo CME, but the measured velocity values are the least accurate in this situation. The velocity of halo CMEs is found to be $600-700 \text{ km s}^{-1}$, more than 50% lower than the actual CME speed.

The heliographic longitude of the CME is fixed at $\varphi_{\text{CME}} = 180^{\circ}$ and when approximating the CME as a single point, the projected velocity can be estimated as

$$v_{\rm proj} = v_{\rm CME} \times \sin \varphi_{\rm obs}.$$

With $v_{\text{CME}} = 1594 \text{ km s}^{-1}$ and $\varphi_{\text{obs}} = 130^{\circ}$ the expected projected velocity, v_{proj} , in the principal direction of the CME (corresponding to $\alpha = 270^{\circ}$), would be 1221 km s⁻¹. The CME velocity measured by CACTus for that configuration is around 1300 km s⁻¹, which is a reasonably good approximation.

Secondly, we changed not only the heliographic longitude, but as well the co-latitude of the observer:

$$0^{\circ} \le \theta_{\rm obs} \le 90^{\circ},$$
$$90^{\circ} \le \varphi_{\rm obs} \le 180^{\circ}.$$

Here as well, the results confirm the complexity of the projection effects on the velocity measurements. However, in this case the projection effects on the velocity measurements are smaller and moreover the value of detected principal angle gives more information about the actual propagation direction of the CME.

Supposing again that the CME is one single particle, with $0^{\circ} \le \theta_{obs} \le 90^{\circ}$ and $90^{\circ} \le \varphi_{obs} \le 180^{\circ}$, we expect a detected velocity value of

$$v_{\text{proj}} = v_{\text{CME}} \times \sin | \arccos(-\sin \theta_{\text{obs}} \times \cos \varphi_{\text{obs}}) |.$$

Table 1 Measured and states					
$(v_{\rm m})$ in the principal direction of	$\theta_{\rm obs}$	$\varphi_{\rm obs}$	$v_{\rm CME}$	$v_{\rm proj}$	$v_{ m m}$
the CME ($\alpha = 270^{\circ}$) in comparison with the expected	90°	130°	1594 km s ⁻¹	1221 km s ⁻¹	1300 km s ⁻¹
values when taking the projection	<i>c</i> 00	1.400	15041 -1	11001 -1	1000 km s
effects into account (<i>v</i> _{max}) and	60°	140°	1594 km s	1192 km s ⁻¹	1200 km s 1
the actual CME velocity (v_{CME}).	60°	150°	1594 km s^{-1}	1054 km s^{-1}	1150 km s^{-1}
('CME)	60°	160°	1594 km s^{-1}	$926 \mathrm{km s^{-1}}$	$1050 {\rm km s^{-1}}$

Table 1 gives the measured velocity values in the principal direction of the CME, in comparison with the expected values taking the projection effects into account.

CACTus detects higher velocity values than those estimated in the simplified approach of a CME being a point. Because of the cloud character of the CME and the Thomson scattering of the electrons, approximating the CME as a single point is an oversimplification. Reconstructing the true velocity from the detected value remains an ill-posed problem. The shape of the 3D cloud has an obvious influence on the appearance of the CME in the whitelight images. These complex characteristics of a solar event hinder the exact calculation of velocity based on (simulated) observations. The cloud character of the simulated CME was quantified by measuring the CME velocity in different directions (different meridional planes); see Table 2 in the Appendix.

Methods for *e.g.* estimating the radial position of a CME also often approximate a CME as a point. Lugaz *et al.* (2010) also confirm that the consequences of such a simplification should not be underestimated in the context of the projection effects and the velocity measurement.

6. Conclusions and Outlook

In this paper we have used simulated CME data as an ideal, objective reference to interpret the automated CME analysis embedded in the CACTus software. Line-of-sight projection and Thompson scattering were applied on the data coming from a 3D MHD simulation (see the Appendix) so as to produce coronagraph-like images. The response of CACTus on these simulated CME images was studied. The advantage of using simulated CME data is that we have prior knowledge of the CME parameters and that we can change the CME-observer perspective at free will.

By comparing the *a priori* known CME characteristics with the characteristics detected with CACTus, we reached the following main conclusions.

The detection scheme of CACTus is sensitive to the "context" in which the CME is present. The simulated CME was presented to CACTus as a limited number of images corresponding just to the lifetime of the event. In addition, the simulated images are free of any noise (instrumental or solar). In contrast, CACTus was developed for essentially unlimited sequences of coronagraph images with high noise levels. At first, the unusual context of the simulated CME made CACTus fail to detect the CME, and caused erroneous velocity measurements as well as a biased classification of the event. Small modifications (*e.g.* inserting non-event images prior to the CME, changing absolute to relative thresholds) allowed to overcome this problems. Nevertheless, the lesson learned is that one must take care when applying CACTus in different circumstances (*e.g.* solar minimum *versus* solar maximum).

After the above-mentioned technical adaptations, CACTus is capable of detecting the CME and its main characteristics with reasonable accuracy.

For a variable CME-observer perspective, CACTus recovers the expected projected speed and projected principal angle successfully. Nevertheless the extended "cloud character" of a CME significantly influences the measurement of the CME velocity and the principal angle. As a consequence, even 3D observational data (detections of the same CME seen from different angles/detectors) do not make it straightforward to calculate the main characteristics of a CME. Our results give an idea of the error margins on the measured speed and angles.

Also a few minor technical issues were identified. For example, CACTus loses the symmetry of the initially symmetric, simulated data. Apparently, the image manipulation introduces numerical errors that can have noticeable effect on the end result. A deeper analysis on this point might be necessary.

The insight uncovered in this paper will guide us when generalizing CACTus to new datasets. They might be other coronagraphs such as the upcoming PROBA3 mission of ESA or EUV images with a relatively extended field-of-view such as STEREO/EUVI or PROBA2/SWAP (Berghmans *et al.*, 2006; Halain *et al.*, 2010). In the latter case we must take into account that CMEs can show significant acceleration close to the solar surface and hence the Hough transform in CACTus might need to be extended to the generalized Hough transform for detecting parabolic propagation paths.

Acknowledgements The development of CACTus as a space weather tool was finalized under ESA contract 16913/03/NL/LvH in the ESA-Space Weather Applications Pilot Project (SWAP). This research was funded by projects GOA/2009-009 (K.U.Leuven), G.0304.07 (FWO-Vlaanderen), 3E090665 (FWO-Vlaanderen), and C 90347 (ESA Prodex 9). Financial support by the European Commission through the SOLAIRE Network (MTRN-CT-2006-035484) and support from the European Commission's Seventh Framework Programme (FP7/2007-2013) under the grant nr. 218816 (SOTERIA project) is gratefully acknowledged. The numerical results were obtained on the HPC cluster VIC of the K.U.Leuven. K. Bonte would like to acknowledge the warm hospitality she received and the insightful discussions she enjoyed during her three week visit at NRL, Washington. She also wants to thank Bram Bourgoignie for a useful co-operation on CACTus code programming.

Appendix: Simulation of a CME Event

For the 3D CME simulations, the ideal MHD equations were solved in spherical coordinates (r, θ, φ) on a three-dimensional spherical mesh, covering a complete sphere, *i.e.* $\theta \in [0, \pi]$ and $\varphi \in [0, 2\pi]$. The magnetic field is kept divergence-free by using the alternative formulation of the constraint transport method (Evans and Hawley, 1988) in terms of the vector potential. The computational domain covered the region between the lower corona and 220 R_{\odot} , using a grid resolution of $1100 \times 91 \times 180$ cells. The grid shows an accumulation of cells both toward the solar surface and toward the solar equator, where the grid size varies from $\Delta r = 0.02 R_{\odot}$ near the solar surface to $\Delta r = 0.25 R_{\odot}$ at $30 R_{\odot}$ and from $\Delta \theta = 4^{\circ}$ near the poles to $\Delta \theta = 0.8^{\circ}$ at the equator. The grid was taken equidistant in the azimuthal direction. The versatile advection code (VAC: Tóth, 1996) was used to integrate the ideal MHD equations over time. The VAC code is a structured finite volume code, running in parallel on distributed memory machines. The simulation up to 1 AU lasted until a time of t = 100 h was reached. For this simulation 440 processors were used and the total duration of the run was of the order of 10 days.

To construct the solar wind, the MHD equations are solved in a co-rotating frame along with an extra-added gravitational force as well as an additional heating source term, very similar to what was used by Manchester *et al.* (2004). This wind model shows no dependence on the azimuthal direction and provides a good approximation for the nearly axisymmetric wind occurring at solar minimum. The ambient magnetic field of the Sun is a simple

dipole with strength of 2.2 gauss (G) at the poles. The solar wind is relaxed in time until a stationary solution is obtained. On top of the stationary solar wind a twisted flux rope structure is placed, with foot points anchored in the solar surface. The solution of the magnetic field in the flux rope is the modified Miller–Turner solution for force free fields inside a torus, as presented by Romashets and Vandas (2003). The flux rope is subjected to the transformation formulas described in Gibson and Low (1998), deforming the flux rope and stretching it toward the solar surface. The transformation allows a deformation of the original flux rope into an expanding shape, imitating a rising prominence. Prominences are always observed over regions where the magnetic field changes sign. Since in the background coronal model for solar minimum the only polarity inversion line coincides with the equator, the modified Miller-Turner solution is placed over the solar equator. From observations it is known that filaments lie in a highly sheared fashion over the inversion line, making an angle of $\approx 20^{\circ}$ (Leroy, Bommier, and Sahal-Brechot, 1984). The positive side of the flux rope is located in the northern hemisphere, where the background magnetic field has a positive sign, and the negative side is located in the southern hemisphere and the center line of the flux rope makes an angle of 20° with the solar equator. The magnetic field strength in a quiescent prominence is typically between 5 and 40 G. In the model, the value for the toroidal field in the center of the flux rope was set to $B_0 = 1.44$ G. This is lower than what is observed, but remember that the initial condition represents a prominence that is already erupting. In order to trigger a CME, the flux rope is given an additional density and radial velocity. The total mass inside the flux rope is 4.3×10^{15} g and the peak velocity inside the flux rope was set to $v_{\rm CME} = 4000 \text{ km s}^{-1}$. The total amount of energy added to the solar wind by the inclusion of the flux rope was 4.31×10^{31} erg, of which 3×10^{30} erg magnetic energy and 4.01×10^{31} erg being kinetic energy. Gopalswamy *et al.* (2005) studied the arrival times of several historical fast events and they argued that the maximum initial speed of a CME may not be much higher than ≈ 3000 km s⁻¹. In the present simulation, the plasma blob was given an initial speed in order to mimic the eruption. Since no initiation mechanism was considered, the mechanism for accelerating the CME is not captured well in this simulation and the plasma cloud will experience a strong deceleration in the initial stages. Therefore, the $v_{\rm CME}$ parameter is set to a quite high velocity of 4000 km s⁻¹. However, this high velocity



Figure 15 The initial state seen from two different view points. The solar surface is color-coded with the radial magnetic field strength. Inside the flux rope the isosurface of density $\rho = 2\rho_*$ is plotted, with ρ_* the surface density. The isosurface is color-coded with the radial velocity.



Figure 16 Visualizing the simulated CME: Three snapshots showing the magnetic field lines from two different view points. The solar surface is color-coded with the radial magnetic field strength. The isosurface represents the location where the density is twice the original background solar wind density, and is colorcoded with the radial velocity.

Table 2 Average velocity and acceleration of the CME front within $30 R_{\odot}$ in three different meridional planes.		$\varphi = 150^{\circ}$	$\varphi = 180^{\circ}$	$\varphi = 210^{\circ}$
	$\langle v \rangle$ $\langle a \rangle$	1059 km s^{-1} -21.3 m s ⁻²	1594 km s ⁻¹ -78.8 m s ⁻²	931 km s ⁻¹ -29.5 m s ⁻²

is only reached in one point of the flux rope and the average amount of extra velocity added corresponds to a value of only $\approx 600 \text{ km s}^{-1}$. The initial state is illustrated in Figure 15.

Figure 16 shows a 3D visualization of three snapshots with a time difference between each of one hour. Remark that the propagation direction of the CME coincides with the negative X-axis ($\theta = \pi/2, \varphi = \pi$). From Figure 16 it may be noted that the CME front is not of circular shape, and the CME propagates the fasted along the negative X-axis. When cutting the CME with different meridional planes ($\varphi = 150^\circ$, 180°, and 210°), the average velocity and acceleration of the CME within 30 R_{\odot} can be found by fitting the height-time curves with second order polynomials. The results of those fits are listed in Table 2. The extensive statistical study of Yashiro *et al.* (2004) pointed out that, on average, the acceleration of CMEs with an average velocity $\langle v \rangle \ge 900 \text{ km s}^{-1}$ is -15 m s^{-2} . The strong deceleration of 79 m s⁻² of the CME front along the $\varphi = 180^{\circ}$ -plane is rare, but not impossible.

References

- Aschwanden, M.J., Wuelser, J.P., Nitta, N.V., Lemen, J.R.: 2009, Solar Phys. 256, 3. doi:10.1007/ s11207-009-9347-4.
- Ballester, P.: 1994, Astron. Astrophys. 286, 1011.
- Berghmans, D.: 2002, In: Kuijpers, J. (ed.) Solar Variability: From Core to Outer Frontiers, ESA SP-506, 85.
- Berghmans, D., Hochedez, J.F., Defise, J.M., Lecat, J.H., Nicula, B., Slemzin, V., et al.: 2006, Adv. Space Res. 38, 1807. doi:10.1016/j.asr.2005.03.070.
- Boursier, Y., Lamy, P., Llebaria, A.: 2009, Solar Phys. 256, 131. doi:10.1007/s11207-009-9358-1.
- Ciaravella, A., Raymond, J.C., Kahler, S.W.: 2006, Astrophys. J. 652, 774. doi:10.1086/507171.
- Cohen, O., Attrill, G.D.R., Manchester, W.B., Wills-Davey, M.J.: 2009, Astrophys. J. 705, 587. doi:10.1088/ 0004-637X/705/1/587.
- Dal Lago, A., Vieira, L.E.A., Echer, E., Gonzalez, W.D., de Gonzalez, A.L.C., Guarnieri, F.L., Schuch, N.J., Schwenn, R.: 2004, *Solar Phys.* 222, 323. doi:10.1023/B:SOLA.0000043566.21049.82.
- Domingo, V., Fleck, B., Poland, A.I.: 1995, Solar Phys. 162, 1. doi:10.1007/BF00733425.
- Evans, C.R., Hawley, J.F.: 1988, Astrophys. J. 332, 659. doi:10.1086/166684.
- Gibson, S.E., Low, B.C.: 1998, Astrophys. J. 493, 460. doi:10.1086/305107.
- Gopalswamy, N., Yashiro, S., Liu, Y., Michalek, G., Vourlidas, A., Kaiser, M.L., Howard, R.A.: 2005, J. Geophys. Res. 110, 09. doi:10.1029/2004JA010958.
- Halain, J., Berghmans, D., Defise, J., Renotte, E., Thibert, T., Mazy, E., et al.: 2010, In: Armand, M., Murray, S.S., Takahashi, T. (eds.) Space Telescopes and Instrumentation 2010: Ultraviolet to Gamma Ray, Proc. SPIE 7732, 77320R. doi:10.1117/12.857979.
- Howard, R.A., Sheeley, N.R. Jr., Michels, D.J., Koomen, M.J.: 1985, J. Geophys. Res. 90, 8173. doi:10.1029/ JA090iA09p08173.
- Hundhausen, A.J., Burkepile, J.T., St. Cyr, O.C.: 1994, J. Geophys. Res. 99, 6543. doi:10.1029/93JA03586.
- Jacobs, C., Roussev, I.I., Lugaz, N., Poedts, S.: 2009, Astrophys. J. Lett. 695, 171. doi:10.1088/0004-637X/ 695/2/L171.
- Jahne, B.: 1997, Practical Handbook on Image Processing for Scientific Applications. CRC Press, Boca Raton, 483.
- Kaiser, M.L., Kucera, T.A., Davila, J.M., St. Cyr, O.C., Guhathakurta, M., Christian, E.: 2008, Space Sci. Rev. 136, 5. doi:10.1007/s11214-007-9277-0.
- Leroy, J.L., Bommier, V., Sahal-Brechot, S.: 1984, Astron. Astrophys. 131, 33.
- Liewer, P.C., Dejong, E.M., Hall, J.R., Pournaghshband, V.J., Thernisien, A., Howard, R.: 2006, AGU Fall Meeting Abstracts, A1472+.
- Lugaz, N., Vourlidas, A., Roussev, I.I., Morgan, H.: 2009, Solar Phys. 256, 269. doi:10.1007/ s11207-009-9339-4.
- Lugaz, N., Hernandez-Charpak, J.N., Roussev, I.I., Davis, C.J., Vourlidas, A., Davies, J.A.: 2010, Astrophys. J. 715, 493. doi:10.1088/0004-637X/715/1/493.
- Manchester, W.B., Gombosi, T.I., Roussev, I., Ridley, A., De Zeeuw, D.L., Sokolov, I.V., Powell, K.G., Tóth, G.: 2004, J. Geophys. Res. 109, 02107. doi:10.1029/2003JA010150.
- Michałek, G., Gopalswamy, N., Lara, A., Manoharan, P.K.: 2004, Astron. Astrophys. 423, 729. doi:10.1051/0004-6361:20047184.
- Mierla, M., Inhester, B., Marqué, C., Rodriguez, L., Gissot, S., Zhukov, A.N., Berghmans, D., Davila, J.: 2009, Solar Phys. 259, 123. doi:10.1007/s11207-009-9416-8.
- Minnaert, M.: 1930, Z. Astrophys. 1, 209.
- Poedts, S., Jacobs, C., van der Holst, B., Chané, E., Keppens, R.: 2009, Earth Planets Space 61, 599.
- Ponz, J.D., Thompson, R.W., Munoz, J.R.: 1994, Astron. Astrophys. Suppl. Ser. 105, 53.
- Robbrecht, E., Berghmans, D.: 2004, Astron. Astrophys. 425, 1097. doi:10.1051/0004-6361:20041302.
- Robbrecht, E., Berghmans, D., Van der Linden, R.A.M.: 2009, Astrophys. J. 691, 1222. doi:10.1088/ 0004-637X/691/2/1222.
- Romashets, E.P., Vandas, M.: 2003, In: Wilson, A. (ed.) Solar Variability as an Input to the Earth's Environment, ESA SP-535, 535.
- St. Cyr, O.C., Plunkett, S.P., Michels, D.J., Paswaters, S.E., Koomen, M.J., Simnett, G.M., et al.: 2000, J. Geophys. Res. 105, 18169. doi:10.1029/1999JA000381.
- Thernisien, A.F.R., Howard, R.A., Vourlidas, A.: 2006, Astrophys. J. 652, 763. doi:10.1086/508254. Tóth, G.: 1996, Astrophys. Lett. Commun. 34, 245.

Tousey, R.: 1973, In: Rycroft, M.J., Runcorn, S.K. (eds.) Space Research XIII, 713.

- Vršnak, B., Sudar, D., Ruždjak, D., Žic, T.: 2007, Astron. Astrophys. 469, 339. doi:10.1051/0004-6361:20077175.
- Wang, Y.-M., Sheeley, N.R. Jr., Howard, R.A., Kraemer, J.R., Rich, N.B., Andrews, M.D., Brueckner, G.E., et al.: 1997, Astrophys. J. 485, 875. doi:10.1086/304467.
- Webb, D.F., Mizuno, D.R., Buffington, A., Cooke, M.P., Eyles, C.J., Fry, C.D., et al.: 2006, J. Geophys. Res. 111, 12101. doi:10.1029/2006JA011655.
- Yashiro, S., Gopalswamy, N., Michalek, G., St. Cyr, O.C., Plunkett, S.P., Rich, N.B., Howard, R.A.: 2004, J. Geophys. Res. 109, 07105. doi:10.1029/2003JA010282.