Available online at:
www.swsc-journal.org

Flares, coronal mass ejections and solar energetic particles and their space weather impacts

RESEARCH ARTICLE

OPEN ⊘ ACCESS

A probabilistic approach to the drag-based model

Gianluca Napoletano[1], Roberta Forte[2], Dario Del Moro[2,*], Ermanno Pietropaolo[1], Luca Giovannelli[2] and Francesco Berrilli[2]

[1] Dipartimento di Scienze Fisiche e Chimiche, Università degli studi dell'Aquila, Via Vetoio snc, 67100 Coppito (AQ), Italy
[2] Dipartimento di Fisica, Università degli studi di Roma "Tor Vergata", Via della Ricerca Scientifica 1, 00133 Rome, Italy

Abstract – The forecast of the time of arrival (ToA) of a coronal mass ejection (CME) to Earth is of critical importance for our high-technology society and for any future manned exploration of the Solar System. As critical as the forecast accuracy is the knowledge of its precision, i.e. the error associated to the estimate. We propose a statistical approach for the computation of the ToA using the drag-based model by introducing the probability distributions, rather than exact values, as input parameters, thus allowing the evaluation of the uncertainty on the forecast. We test this approach using a set of CMEs whose transit times are known, and obtain extremely promising results: the average value of the absolute differences between measure and forecast is 9.1h, and half of these residuals are within the estimated errors. These results suggest that this approach deserves further investigation. We are working to realize a real-time implementation which ingests the outputs of automated CME tracking algorithms as inputs to create a database of events useful for a further validation of the approach.

Keywords: Heliosphere / coronal mass ejection (CME) / space weather

1 Introduction

Coronal mass ejections (CMEs) are violent phenomena of solar activity with repercussions throughout the entire heliosphere. Their manifestations into interplanetary space are responsible for major geomagnetic storms, hence the prediction of the arrival of interplanetary coronal mass ejections (ICMEs) at 1 AU is one of the primary subjects of the space-weather forecasting (e.g. Daglis, 2001; Schrijver and Siscoe, 2010).

Several forecasting methods have been proposed over the last two decades. On one hand, there are the approaches relying on statistical-empirical relations established through observations between coronagraphically measured parameters and quantities related to their heliospheric propagation (e.g. Brueckner et al., 1998). Another approach is represented by numerical MHD-based models of the heliospheric propagation of ICME, generally requiring a detailed knowledge of the state of the heliosphere and large computational facilities (as is the case for the WSA-ENLIL model Odstrcil and Pizzo, 1999; Odstrcil et al., 2004; Owens et al., 2005; Parsons et al., 2011). The numerical models are fairly accurate (Vršnak et al., 2014), and highly sensitive to the quality of the input parameters (Falkenberg et al., 2010b), as one may expect. Recently, the WSA-ENLIL model started to be

employed also in a probabilistic approach (Cash et al., 2015; Mays et al., 2015; Pizzo et al., 2015) to quantify the prediction uncertainties and to determine the forecast confidence. However this approach, interesting as it is, is not widely enough used for real-time space-weather forecasting due to the demanding computational needs.

The last category, somewhat lying in between the previous ones, is that employing an MHD- or HD-based simplified description of the interactions an ICME may be subjected to during its interplanetary travel Gopalswamy et al. (2000); Vršnak and Gopalswamy (2002); Michalek et al. (2002); Schwenn et al. (2005). Such an approach leads to analytical models or empirical models which require modest computational power. Those models assume a morphology (either simple as in Möstl et al. (2011) or more complex as in Isavnin (2016)), a fixed direction and a velocity evolution for the CME and can predict an arrival time and speed from relatively limited initial information on the CME onset conditions. Such initial conditions can be obtained from several sources, such as LASCO C2 and C3 coronal imagers on-board the SOlar and Heliospheric Observatory (SOHO Domingo et al., 1995) and, more recently, from COR1 and COR2 and the Heliospheric imager (HI) on-board the Solar Terrestrial Relation Observatory (STEREO Kaiser et al., 2008) spacecraft either separately or using appropriate tools to merge the measures taken from the different instruments and the different points of view (Lugaz et al., 2009; Davies et al., 2012; Möstl and Davies, 2013;

---

*Corresponding author: dario.delmoro@roma2.infn.it

Möstl et al., 2014). In this paper, we focus our attention on a model belonging to the last category, the drag-based model (DBM Vršnak et al., 2013). The model hypothesizes a simple interaction between the ICME plasma and the solar wind that works to equalize the ICME velocity to that of the solar wind itself. This is consistent with the measures of the ICME speeds in the near Earth environment which are typically confined in the 400–700 km/s range and the estimates of the initial velocity of the plasma ejecta near the Sun, which range between 100 km/s and 2000 km/s. This process has been modeled analogously as an aerodynamic or viscous drag by several authors (Cargill et al., 1995; Vršnak and Gopalswamy, 2002; Shi et al., 2015). It makes use of the initial CME velocity, its distance from the Sun at the moment of the measure, and the solar wind speed to compute the travel time at 1 AU.

The DBM has already generated a whole family of approaches, which may differ for the way to evaluate the initial parameters, or how the CME is propagated in the heliosphere.

The difference may arise from the peculiarity of the data used (type and source) and from the interpretation of such data to estimate the CME onset parameters, which ultimately depends on the shape assumed for the CME itself (the Fixed-method in Sheeley et al. (1999); Rouillard et al. (2008); Möstl et al. (2014), the Harmonic mean method in Howard and Tappin (2009); Lugaz et al. (2009), the Self Similar Expansion method in Davies et al. (2012); Möstl and Davies (2013), the graduated cylindrical shell method (GCS) in Thernisien et al. (2006, 2009), the Elliptic Conversion method in Rollett et al. (2016), to cite the most used). Or, the difference can be in the way the drag effect is approximated and the velocity of the CME evolves as in Hess and Zhang (2015), Žic et al. (2015) and Rollett et al. (2016).

In the cited literature, much attention is paid to get the best estimate of the DBM parameters and an evaluation of the associated errors, but none of the mentioned DBM approaches takes into consideration this last information in the implementation of the forecast.

In this work, we apply a statistical approach on the DBM for the computation of ICME travel times, by introducing the probability distributions rather than exact values for the input parameters. This approach has the non-trivial advantage to provide also an evaluation of the uncertainty on the arrival time. In Section 2 we rapidly revisit the equations of the DBM model and introduce its probabilistic version. We also present and discuss the probability distribution functions (PDFs) that we assume to compute the most probable ICME travel times. Section 3 presents the dataset of CME speeds, onset times and travel times that we use to compute and then compare (Sect. 4) the forecast travel times and associated errors. In Section 5 we comment on our results, comparing with those already present in the literature, and discuss possible applications and further evolutions of this model.

For the sake of clarity, we specify that in this work the terms CME and ICME are referred to the plasma and magnetic field structure expelled from the Sun, without the shock that precedes it.

2 The drag-based model

2.1 General description

The drag-based model relies on the hypothesis that all the interactions responsible for the launch of the CME cease in the upper corona, and that, beyond a certain distance, the dynamics of ICME propagation are governed mainly by its interaction with the ambient solar wind. The DBM considers such an interaction by means of a drag force analogous to that experienced by a body immersed in a fluid. The idea of an MHD analogous "hydrodynamical" drag is supported by the observation that ICMEs which are faster than the solar wind are decelerated, whereas those slower than the solar wind are accelerated by the ambient flow (Gopalswamy et al., 2000; Manoharan, 2006).

Following Cargill (2004), we consider the relative speed dependence of the drag force in the radial direction:

$$F_d = -C_d A \rho (v - w)|v - w|, \tag{1}$$

where v is the ICME radial speed and w that of the solar wind, A is the ICME cross-section, ρ is the solar wind density and C_d is a dimensionless coefficient for the drag force. In a classical Newton's law framework, this leads to a radial drag acceleration in the form:

$$a = -\gamma(v - w)|v - w|, \tag{2}$$

where γ is the so-called drag parameter which contains the information about the ICME shape, mass, and in general about the effectiveness of the drag effect.

Considering the solar wind speed and the drag parameter as constants (which is a good approximation beyond $20 - 40\ R_\Theta$ Cargill, 2004; Vršnak et al., 2013), equation (2) can be solved explicitly, obtaining as functions of time the ICME speed:

$$v(t) = \frac{v_0 - w}{1 \pm \gamma(v_0 - w)t} + w, \tag{3}$$

and the heliospheric distance:

$$r(t) = \pm \frac{1}{\gamma} \ln[1 \pm \gamma(v_0 - w)t] + wt + r_0, \tag{4}$$

where the $\pm$ signs apply to the cases $v_0 > w$ and $v_0 < w$, respectively, and r_0 and v_0 are the CME distance from the Sun and velocity at the onset time t_0. In this framework, the model needs four quantities, $[r_0, v_0, w, \gamma]$, to compute the heliospheric distance and velocity of the ICME at any t.

The shape of the ICME we are modeling corresponds to type A) in Figure 9 of Schwenn et al. (2005), i.e. the front of the CME is a section of a sphere concentric with the Sun.

2.2 The probabilistic drag-based model

As just stated, the DBM needs four quantities to be computed, namely $[r_0, v_0, w, \gamma]$. The first two quantities suffer from measure errors, while the last two are, in general, unknown.

If we consider the measure errors to be described by Gaussian PDFs, and assume a priori PDFs for both w and γ, we can extend the DBM into a probabilistic approach.

The Probabilistic drag-based model (P-DBM henceforth), is a Monte-Carlo evaluation of the time of arrival (ToA) and the velocity of the ICME at a chosen distance from the Sun, transforming the PDFs associated to the inputs into PDFs for the outputs, thus generating best estimates and errors for both
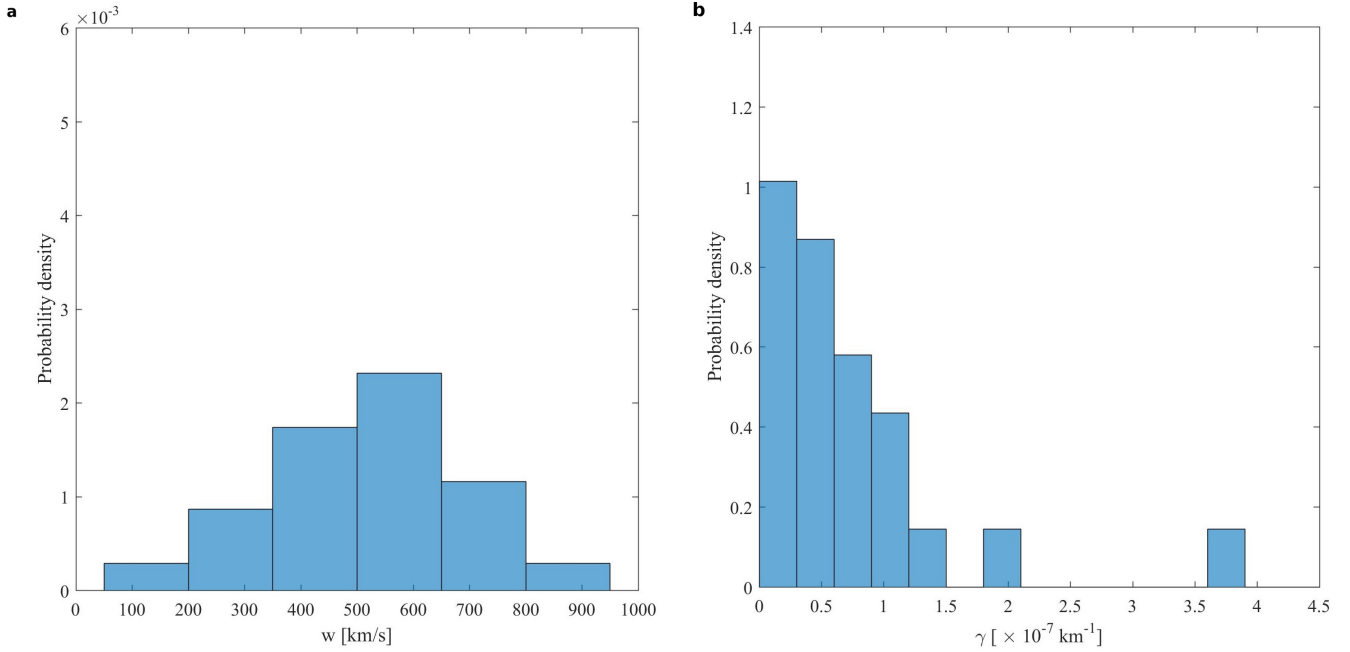
Fig. 1. Histograms of w (a) and γ (b) obtained by the inversion of Schwenn et al. (2005) and Manoharan (2006) catalogs.

the ToA and the velocity. For each ICME whose r_0 and v_0 are measured, we can generate N different $[r_0, v_0, w, \gamma]$ initial conditions sets, randomly chosen from the relative PDFs, to compute via equations (3) and (4) the transit time and the velocity at 1 AU, for example. This process generates the PDFs associated to t_{1AU} and v_{1AU}, which can be used to estimate the ICME most probable ToA and velocity and their associated uncertainties at 1 AU.

Of course, the robustness of the results strongly depends on the validity of the assumptions, the realism of the PDFs, and on a thorough exploration of the parameter space, i.e. how large is N. Given the simplicity of equations (3) and (4) and the present computing capabilities, N of the order of $10^4 - 10^6$ can be used to explore the parameter space and obtain nicely sampled output PDFs in a matter of seconds.

2.3 PDFs for the input quantities

In this section we introduce the PDFs which will be used for the four input quantities.

As Vršnak et al. (2013) have shown, the two equations (3) and (4) can be inverted to obtain the drag parameter γ and the solar wind speed w, if the initial position r_0 and speed v_0 of an ICME and its ToA t_{1AU} and velocity v_{1AU} at 1 AU are known.

$$\gamma = \frac{(v_0 - v_{1AU})}{(v_0 - w)(v_{1AU} - w)t_{1AU}}. \quad (5)$$

This equation can be used to compute directly γ once one has numerically solved:

$$\frac{(v_0 - w)(v_{1AU} - w)t_{1AU}}{(v_0 - v_{1AU})} \ln\left[\frac{(v_0 - v_{1AU})}{(v_{1AU} - w)} + 1\right]$$
$$+ wt_{1AU} + r_0 - r_{1AU} = 0, \quad (6)$$

to obtain w.

As in Vršnak et al. (2013), we use the catalogs of ICMEs by Schwenn et al. (2005) and Manoharan (2006) to compute this inversion. The first list consists of 91 CMEs between 1991 and 2001 for which the authors were able to uniquely associate ICME signatures in front of the Earth after careful inspections of the SOHO/LASCO CME catalog and the complete LASCO/EIT data set. The second list by Manoharan consists of 30 CME events between 1998 and 2004 whose heliospheric evolution has been investigated between the Sun and the Earth using LASCO coronagraphic images and interplanetary scintillation images of the inner heliosphere. Therefore, these lists include CME events for which a safe association between a remote coronagraphical observation and an in situ signature has been established, allowing the knowledge of quantities such as transit time and initial and final speed, required for the inversion. From the results, we obtain the histograms reported in Figure 1 for w and γ. We choose larger bins than in the original work to make the obtained distributions more robust at the expense of sampling. Apart from those differences, these distributions are of course consistent with Vršnak et al. (2013) results.

For w, we can complement the distribution obtained by using the values of the solar wind recorded by SOHO, ACE, ULYSSES, HELIOS (Schwenn, 1983; Ipavich et al., 1998; Stone et al., 1998; Coplan et al., 2001; McComas et al., 2003; Ebert et al., 2009) and many other missions.

The common understanding (see Schwenn, 2006, for a review) is that there exist two different PDFs for the so-called slow (below 500 km/s) and fast solar wind, the latter originated from the coronal holes, which are regions on the Sun with depressed UV emission and low magnetic activity. The w probability densities that we assume are plotted in Figure 2a, with the slow w represented by a Gaussian PDF centered at 400 km/s with $\sigma = 33$ km/s, and the fast w represented by a Gaussian PDF centered at 600 km/s with $\sigma = 66$ km/s. Of
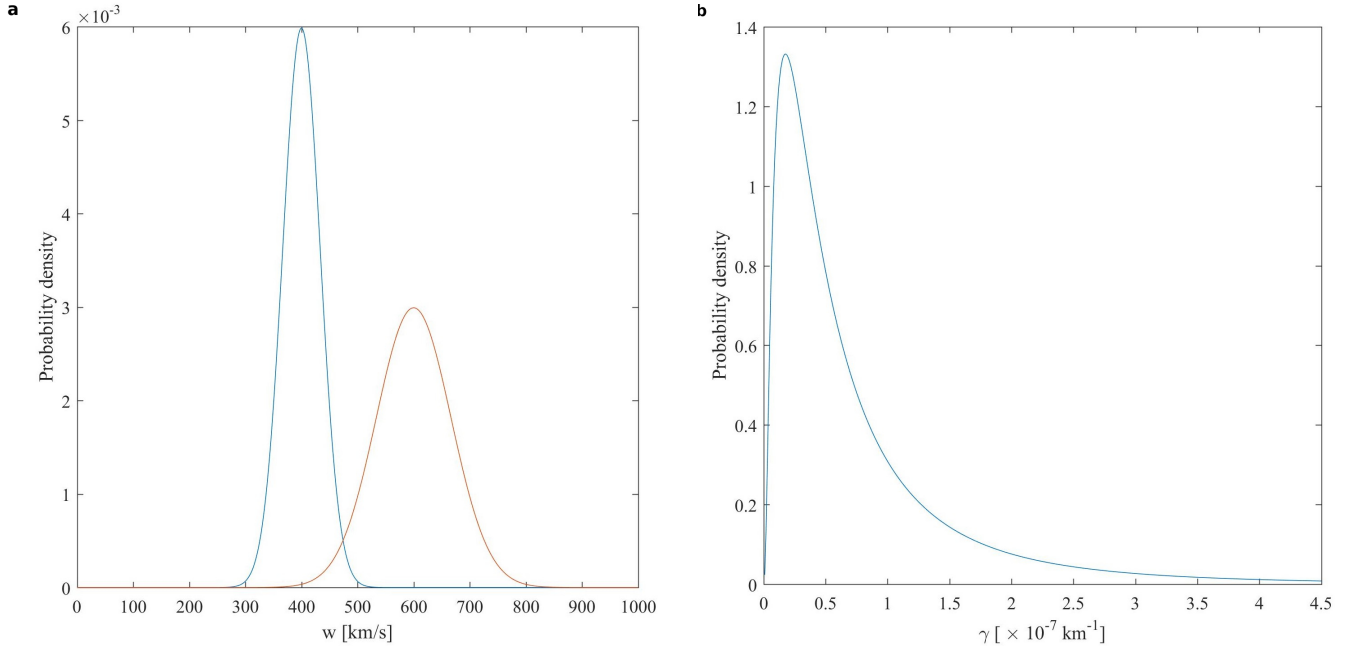
Fig. 2. (a) PDF adopted for the random generation of *w* in the P-DBM, with the slow *w* represented by a Gaussian PDF centered at 400 km/s with $\sigma = 33$ km/s, and the fast *w* represented by a Gaussian PDF centered at 600 km/s with $\sigma = 66$ km/s. (b) PDF adopted for the random generation of *γ* in the P-DBM, modeled by a Log-Normal function with $\mu = -0.70$ and $\sigma = 1.01$.

course, such PDFs are limited to positive values of *w*. Following the works of Robbins et al. (2006); Vršnak et al. (2007), we adopt the fast *w* PDF in those cases where there is a prominent coronal hole in the center on the disk, the slow *w* PDF in all the other cases.

For *γ*, we note that the the distribution retrieved by the inversion has a peak in the first bin $(0.2 - 0.4 \times 10^{-7}\,\mathrm{km}^{-1})$ and then decays, with an extended tail up to $\simeq 4 \times 10^{-7}\,\mathrm{km}^{-1}$. The skewed shape of the distribution suggested to fit $(\tilde{\chi}^2 = 1.13)$ such a distribution with the 2-parameter Log-Normal function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \tag{7}$$

retrieving $\mu = -0.70$ and $\sigma = 1.01$, to obtain an analytic form for the PDF, which is shown in Figure 2b.

Despite the fact that we are not putting forward any physical model for the CME kinematics, we must note that the Log-Normal distribution has been found to describe several aspect of solar wind plasma (see Burlaga and Lazarus, 2000, and references therein) and even the CME speed distribution (Yurchyshyn et al., 2005). In our case, the Log-Normal distribution just provides a good fit to the observed distribution, capturing its properties in just two parameters.

For r_0, we consider that CME detection algorithms have inherent uncertainties for the CME location and the moment and duration of the CME liftoff. From that, we assume that the PDF of r_0 can be modeled by a Gaussian PDF whose average is the last height derived by the CME tracking algorithm at the onset time and whose sigma is estimated from the associated error ($3\sigma \simeq 1R_\odot$ in the case of Shi et al., 2015).

Also for v_0, we assume a Gaussian PDF whose average value is the velocity measured by the CME tracking algorithm and whose sigma is the uncertainty associated to the measurement.

2.4 P-DBM step-by-step

To resume, here is a step-by-step description of how the P-DBM performs a prediction on the arrival of an ICME:

1. the position PDF is generated using the last measured CME height within coronagraph images and its associated error;
2. the velocity PDF is generated using the measured velocity and its associated error;
3. the Log-Normal PDF described by equation (7) ($\mu = -0.70$ and $\sigma = 1.01$) is considered for the drag parameter;
4. a Gaussian PDF is chosen for the solar wind velocity, selecting either fast solar wind conditions (600 ± 66 km/s) in the case of a coronal hole in a relevant position of the solar disk, or slow solar wind otherwise (400 ± 33 km/s);
5. *N* initial condition sets $[r_0, v_0, \gamma, w]$ are randomly generated from those PDFs;
6. *N* different ToAs at 1 AU t_{1AU} are computed from equation (4), by setting $t = t_{1AU}$ and $r(t_{1AU}) = 1$ AU, and computing $t = t_{1AU}$ as the root of the equation via an iterative algorithm;
7. the ToA PDF is evaluated from the *N* t_{1AU} values;
8. the best estimate for t_C and its associated error are evaluated as the mean and the root mean square of the ToA PDF;
9. steps 6–8 are also applied to equation (3) to evaluate the best estimate for v_C and its associated error.

Table 1. Sample of events from Shi et al. (2015) employed to test the P-DBM. Columns are in order: CME index number, CME onset date and time (UT) at 2.5 $R_\odot$, CME initial speed with associated uncertainty, CME onset date and time (UT) at $15R_\odot$, arrival date and time (UT) of the ICME at 1A, solar wind (Slow/Fast) during the CME propagation.

#	Onset time @ 2.5 $R_\odot$	v±3σ [km/s]	Onset time @ 15 $R_\odot$	Arrival @ 1 AU	SW
1	12 Dec 2008, 08:37	363 ± 23	12 Dec 2008, 15:16	17 Dec 2008, 02:00	S
2	03 Apr 2010, 09:54	864 ± 7	03 Apr 2010, 12:42	05 Apr 2010, 08:00	F
3	08 Apr 2010, 03:39	512 ± 34	08 Apr 2010, 08:22	12 Apr 2010, 02:00	S
4	16 Jun 2010, 14:39	222 ± 2	17 Jun 2010, 01:32	21 Jun 2010, 08:00	S
5	15 Feb 2011, 02:24	769 ± 12	15 Feb 2011, 05:33	18 Feb 2011, 03:00	S
6	04 Aug 2011, 04:39	1512 ± 90	04 Aug 2011, 06:15	06 Aug 2011, 12:00	F
7	14 Sep 2011, 00:39	505 ± 5	14 Sep 2011, 05:26	17 Sep 2011, 10:00	S
8	19 Jan 2012, 14:54	1299 ± 16	19 Jan 2012, 16:46	23 Jan 2012, 00:00	S
9	13 Mar 2012, 17:39	1616 ± 17	13 Mar 2012, 19:09	15 Mar 2012, 19:00	F
10	19 Apr 2012, 15:39	607 ± 15	19 Apr 2012, 19:38	23 Apr 2012, 12:00	S
11	12 Jul 2012, 16:39	1224 ± 14	12 Jul 2012, 18:37	15 Jul 2012, 06:00	F
12	28 Sep 2012, 00:39	1104 ± 112	28 Sep 2012, 02:50	01 Oct 2012, 12:00	S
13	05 Oct 2012, 03:39	558 ± 21	05 Oct 2012, 07:59	08 Oct 2012, 18:00	S
14	27 Oct 2012, 16:54	340 ± 28	28 Oct 2012, 00:00	01 Nov 2012, 00:00	S

3 The dataset

In order to test the P-DBM described in the previous section, we use a sample of events from Shi et al. (2015). For such events, a reconstruction of the ICME shape and speed has been obtained with the graduated cylindrical shell model (Thernisien et al., 2006, 2009) by means of triangulation of coronagraphic images taken from both STEREO and LASCO. Following Shi et al. (2015), we excluded from the original sample those ICME which probably had interactions with the background magnetic field or other CMEs. We also excluded entry 11 from the original sample which was most probably not correctly associated with the ICME arrival time (cf. Möstl et al., 2014). The details about how the CGS model has been used to fit the CME shapes and to determine the CME initial speeds and heights are reported in the original paper of Shi et al. (2015). Here, we only recall that the authors estimated the errors of their detection and tracking procedure and that the CME speeds were evaluated through linear fits of the height versus time curves, and the associated error is the uncertainty of the linear fitting. This will be used to estimate the width of the Gaussian PDFs associated to the CME position and velocity uncertainties and to update the original onset times of the events in Shi et al. (2015) which are reported in the second column of Table 1.

These onset times are associated to the first detection in the instrument FOV, that is at 2.5 $R_\odot$. In order to employ the DBM in the proper range of heliospheric distances, we choose to move the onset positions at the last useful detection in the instrument FOV at $15R_\odot$. Consequently, we re-evaluated the onset time for each CME by adding a delay of $12.5R_\odot/v_0$, using the velocity v_0 (third column in Tab. 1) obtained by Shi et al. (2015) through the linear fits of the CME positions exactly between $2.5R_\odot$ and $15R_\odot$. The new onset times are reported in the fourth column of Table 1. Consequently, for the purpose of this work, we can assume for each event a normal distribution of the height r_0 at the new onset time, with mean value $<r_0> = 15\,R_\odot$ and standard deviation $\sigma_r = 0.33R_\odot$.

Furthermore, it must be observed that for the events from the paper by Shi et al. (2015) the ICME arrival time is referred to the time of first occurrence of an ICME signature in the near Earth environment, which in most cases is the ToA of the fore-shock. To perform a correct validation of the CME transit time forecast, we want to consider the arrival of the ICME leading edge, instead of that of the shock (see also the discussion in Schwenn et al., 2005; Vršnak et al., 2014). To this purpose, for each event, we checked for the ToA of a plasma driven effect (Magnetic clouds or Ejecta), as reported in the GMU CME/ICME list compiled by Phillip Hess and Jie Zhang (http://solar.gmu.edu/heliophysics/index.php/GMU_CME/ICME_List). In 10 out of 14 cases, we could correct the arrival times. Column five of Table 1 reports the arrival date and time of the CME, taking into account this update.

The last column of Table 1 reports the condition of the solar wind associated with the CME, obtained by the inspection of suitable coronal images and verified by using data recorded by ACE (Stone et al., 1998).

4 Validation of the P-DBM

We apply the probabilistic approach in order to generate the transit time distribution for each event in Table 1. For this run, the number of forecast realizations has been set to $N = 50000$ and it took less than a minute to obtain the results on a desktop PC.

As example, we show in Figure 3a the distribution of the transit times t_i computed by the P-DBM for the first CME of the sample. As a result from the input distributions, the travel times range from 80 to 120 h, with a median value of 103.8 h. The distribution is not symmetric, slightly skewed towards the shorter times. However, it is viable to describe this distribution by its mean value $t_C = 103.1$ h and its root mean square $\sigma = 4.4$ h.

The results for the whole sample, with t_C and σ of the arrival time distributions taken as the measure of the predicted arrival time, are reported in the third column of Table 2. In the
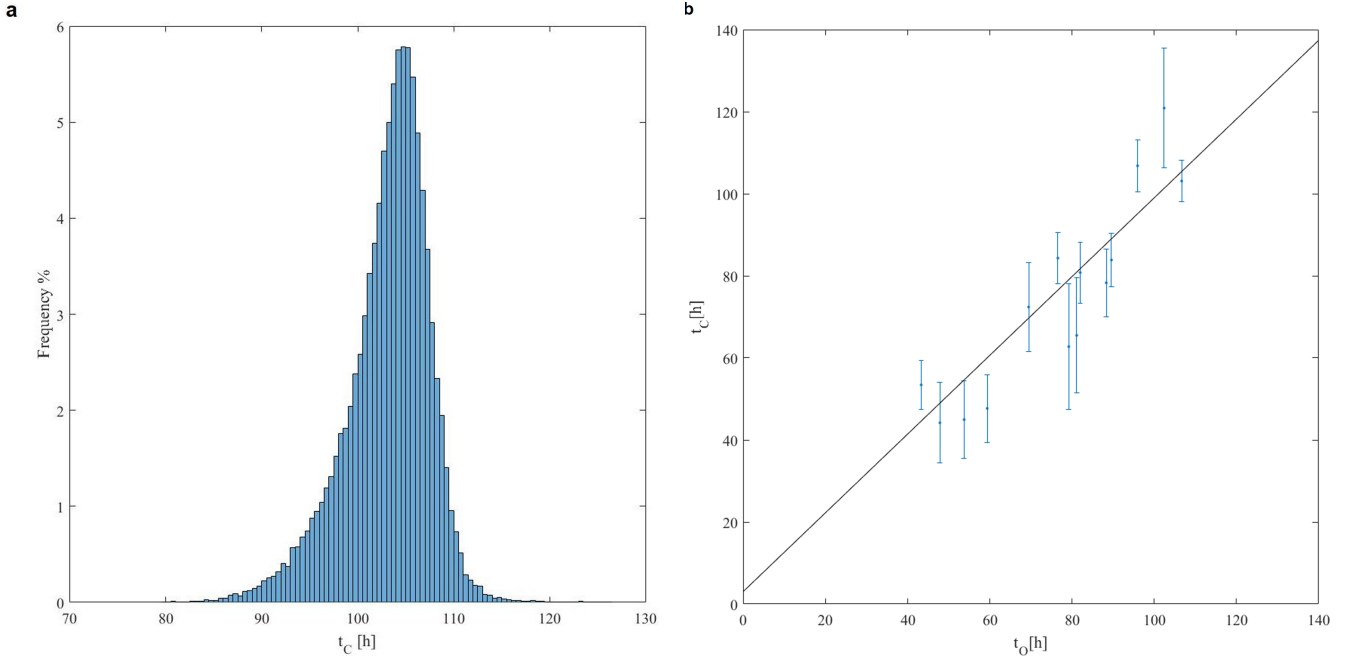
Fig. 3. (a) Distribution of the transit times t_i calculated for event #1 in Table 1. $N = 50000$ initial conditions are generated in the P-DBM. (b) Dots with error bars are the forecast transit times t_C versus observed transit times t_O. The solid line shows a linear fit to the data.

Table 2. Results from the P-DBM statistical simulation for the events in Table 1. In the first column the CME index as in Table 1, in the second column the ICME transit time t_O from $15R_\odot$ to 1 AU, in the third column the computed CME transit time t_C with the associated error σ. In the fourth column the difference $t_O - t_C$.

#	t_O [h]	$<t_C> \pm \sigma$ [h]	$t_O - <t_C>$ [h]
1	106.7	103.1 ± 4.4	3.6
2	43.3	53.4 ± 5.8	-10.1
3	89.6	83.9 ± 6.0	5.8
4	102.5	120.9 ± 14.3	-18.4
5	69.4	72.5 ± 10.7	-2.9
6	53.7	44.9 ± 9.3	8.8
7	76.6	84.3 ± 5.8	-7.7
8	79.2	62.7 ± 15.2	16.5
9	47.8	44.2 ± 9.7	3.7
10	88.4	78.3 ± 8.1	10.1
11	59.4	47.7 ± 8.1	11.7
12	81.2	65.5 ± 14.0	15.7
13	82.0	80.8 ± 7.0	1.2
14	96.0	106.8 ± 5.8	-10.8

second column, instead, we report the observed transit time t_O computed as the difference between the onset time and the arrival time at 1 AU of Table 1.

Figure 3b shows a plot of t_C with 1σ error bars versus t_O and a least squares linear fit to these data. The two datasets are evidently highly correlated, with a correlation coefficient $R = 0.87$. The linear fit performed on the data ($\tilde{\chi}^2 = 1.66$) retrieved a slope of 1.00 ± 0.1 and a constant value of $3\,\text{h} \pm 8\,\text{h}$. Given those values, the P-DBM results are compatible with the $t_C = t_O$ hypothesis.

Similarly to Colaninno et al. (2013), we plot the residuals $t_O - t_C$ and the error associated to t_C for the 14 CMEs in Figure 4a to allow an easy comparison of the forecast results. In particular, for 7 CMEs out of 14 the forecast residuals are within the error. Also, we report in Figure 4b the histogram of the residuals. It can be noted that 80% of the forecasts are within 15 h of the actual t_O, and just one is beyond 20 h. The distribution is compatible with a Gaussian function, centered in zero and with a $\sigma \simeq 10.6\,\text{h}$, with a marginal partiality towards forecasts behind of the observed times. To conclude, we computed the average of the absolute value of the residuals $<|\Delta t|> = 9.1\,\text{h}$, which is often used in the literature to assess the forecast accuracy.

5 Conclusions and future work

In this work, we predicted the transit time between the Sun and the Earth for a sample of 14 CME events. These events were selected among the database of Shi et al. (2015), for which the onset time, initial velocity and transit time are known. By using the DBM (Vršnak et al., 2013) and a probabilistic approach, we were able to associate an error to the transit time we computed, assuming that all the input parameters could be described by suitable PDFs.

For the shape we adopted to model the ICME and since all these ICMEs hit Earth, we did not use the CME principal direction nor the angular width to compute the transit time from the initial parameters. Nevertheless, it is straightforward to modify the P-DBM to include a different CME shape and to consider the PDFs also for those two input parameters. Given the very short time needed to compute a CME transit time distribution with this approach, adding two dimensions to the parameter space to be explored should be still feasible with undemanding computational resources.
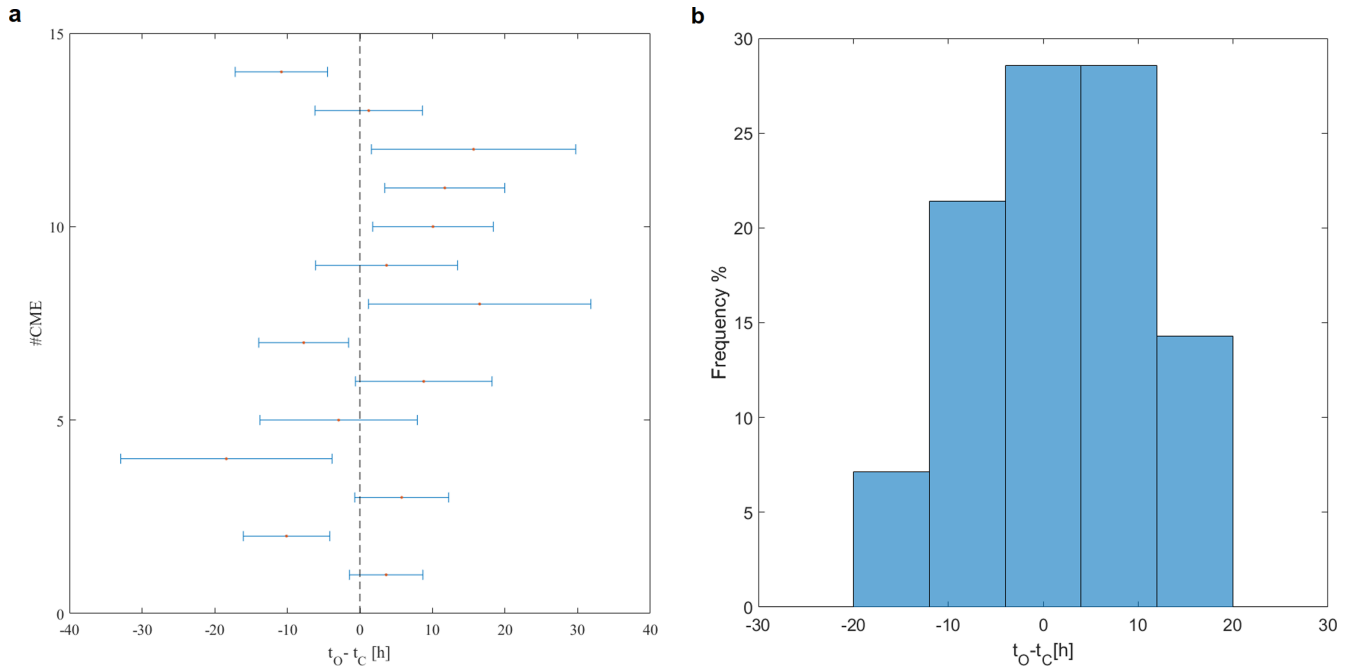
a

b

Fig. 4. (a) The residuals $t_O - t_C$ and the error associated to t_C for the 14 CMEs. (b) Distribution of the residuals $t_O - t_C$.

Even with a model as simple as this, the results of the probabilistic approach are extremely promising:
 – the scatter plot of t_C vs t_O has a slope which is unity within the errors:
 – the histogram of the residuals $\Delta t = t_o - t_C$ has a Gaussian shape, centered in zero and with a $\sigma \simeq 10.6$ h;
 – the average of the absolute value of the residuals is $<|\Delta t|> = 9.1$ h.

However, less than half of the residuals is within the 1σ error associated to t_C, which is under-performing for a Gaussian distribution of the associated error. This can either be due to a statistical fluctuation (given the small dimension of the test set) or to an under-estimate of the input PDF widths. Of course, this disagreement may also arise from the model assumptions. In its simplicity, this DBM implementation models the ICME front as a portion of a sphere concentric with the Sun, therefore neglecting the difference between the ICME apex position and velocity and the ICME position and velocity on the ecliptic plane. On the other hand, this assumption reduces the number of PDFs needed by the model. While it is unclear whether increasing the model complexity will significantly reduce the discrepancy or not, especially considering the intrinsic difficulty in measuring the actual travel times (errors and bias can arise both from the onset and the arrival time estimates), it is instead possible that the PDF we used for γ, evaluated from actual data, may have incorporated most of such complexity, thus including these effects in the model in a statistical way. At present, we can conclude that the chosen PDFs led to good estimations of the average times on transit time forecasts, but we need a larger sample to properly evaluate the robustness of the associated errors. This is probably the main task for future work.

There is a vast literature to compare our results with. We limit ourselves to cases where the authors employed data with projection effects eliminated (measures in quadrature or multi-spacecraft plus CGS model), as in our test

Gopalswamy et al. (2001) found an empirical relation between the initial CME velocity and its acceleration and applied this relation to a model to compute the ToA at Earth. They were able to forecast the ICME ToA at 1 AU with a mean error $<|\Delta t|> = 10.7$ h and 72% of the events had ToA within ± 15 h from the predicted values

Owens and Cargill (2004) tested on a 35 CME sample three different models: a model with a constant acceleration Gopalswamy et al. (2000), a model with an acceleration which ceased before 1 AU Gopalswamy et al. (2001), and the original aerodynamic drag model Vršnak and Gopalswamy (2002). These three model were best fitted on the sample and their $<|\Delta t|>$ varied from 12 to 9 h.

Schwenn et al. (2005) derived an empirical correlation between halo CME expansion speeds and travel times to 1 AU, fitting a straight forward deceleration model assuming viscous drag on the data from 75 halo CME events. For 95% of those events, the shock associated to the CME arrived within ± 24 h of the predicted time.

Colaninno et al. (2013) found that a first-order polynomial to the height-time measurements beyond $50 R_\Theta$ (0.23 AU) was the best parameter for predicting the CME ToA at 1 AU. For a sample of 9 CME, they were are able to predict their ToA to within ± 13 h. It is worth to stress that they supplemented their data with STEREO/HI observations, thus increasing the accuracy of their CME initial parameter estimation.

Taktakishvili et al. (2009) instead evaluated the performances of the ENLIL MHD simulation fed with a cone model of CME for a sample of 14 events. They reported an average absolute error of 6 h, which is also very similar to the

error reported by Millward et al. (2013) of 7.5 h, obtained again with ENLIL simulation initialized with CME parameters obtained via the CME Analysis Tool (CAT), but on a larger (25 events) set. Vršnak et al. (2014) compared the CME arrival time prediction based on the DBM against ENLIL. They reported estimation errors of about 14 h with standard deviation ranges from 14 to 19h, depending on the sample and method.

Shi et al. (2015) used a multi-parametric best fit on the transit time versus the initial speed for different drag based regimes. Depending on the regime, they were able to reach a mean error $<|\Delta t|>$ down to 6.7 h. Since we employed exactly their sample to test the P-DBM, we can note that their model performed better than ours on this sample.

It is worth to stress that Shi et al. (2015) (and all the authors previously cited) fitted their distribution to the data, therefore optimizing the model to that dataset. Our approach, in contrast, used two datasets to build the PDFs and was tested against an independent dataset, thus providing a true a-priori forecast test.

As already stated, we are aware that our results are based on the analysis and comparison of a very limited dataset. Among the next steps in the further validation of this approach, is the test with a larger database of ICME. Since databases which provide information sufficient to fully characterize the ICME are difficult to retrieve, we are already taking into consideration the possibility of having much less information on the ICME onset and morphology.

Therefore, we are working to include both the uncertainty on the angular extension, the uncertainty on the main direction and on the de-projected velocity of the CME in the P-DBM, again, modeled by PDFs. At present, we are also working on a real-time implementation of the P-DBM which ingests the parameters of ICMEs tracked by the CACTUS software (Robbrecht and Berghmans, 2004) and forecast the ToA at 1 AU of the ICMEs and their velocity, of course with the associated errors. As a result, we will build up a database of the results and we plan to verify and possibly re-consider the PDFs we have chosen for the input parameters.

Also, we are pondering an evolution to consider a different morphology of the ICME, passing from the cone model (and its intersection with the ecliptic plane) to a 3-D light-bulb model or a similar model (e.g. Kleimann, 2012).

All these effects significantly alter the travel time and it is worth to explore how the P-DBM can include them in its probabilistic approach.

Since a complete and real-time stereoscopic determination of the CME morphology and propagation will not be available in the near future, all the 3D effects which are not taken into account in the present P-DBM should be considered as partially unknown variables and should be modeled by suitable PDFs, constrained by as much information as available. As example, the real width, direction and velocity of the CME have to be evaluated from images which suffer from projection effects. There are several ways to de-project the data, which imply different assumptions. One of the simplest (Zhao et al., 2002) assumes that the cone-shaped CME has its vertex in the Sun's center and its axis normal to the solar surface at the position of a relevant solar magnetic feature (erupting filament or flaring AR). In such a case, this information, error propagation theory and previous CME parameters statistics could be used to generate the PDFs needed to propagate the

ICME with the P-DBM. It is likely that adding other uncertainties from other input parameters will enlarge the error associated with the forecast, but it is important to stress again that the P-DBM light computation needs make it interesting to evaluate the propagation of any ICME in any portion of the inner solar system.

To conclude, the accurate prediction of the ToA of an ICME to Earth or other interesting part of the Heliosphere (e.g. Falkenberg et al., 2010a) is of critical importance for our high-technology society and for any future manned exploration of the solar system. We think that as critical as the prediction accuracy is the knowledge of precision, i.e. the error associated to the forecast. The method we presented here, building on the DBM model of Vršnak et al. (2013), is capable to predict the arrival time of ICMEs to the Earth and its uncertainty with minor computation necessities, providing a forecast of the space weather in the near Earth environment with a 2-day horizon.

References

Brueckner G, Delaboudiniere J-P, Howard R, Paswaters S, St Cyr O, Schwenn R, Lamy P, Simnett G, Thompson B, Wang D. 1998. Geomagnetic storms caused by coronal mass ejections (CMEs): March 1996 through June 1997. *Geophys Res Lett* **25**: 3019–3022.

Burlaga LF, Lazarus AJ. 2000. Lognormal distributions and spectra of solar wind plasma fluctuations: Wind 1995–1998. *J Geophys Res* **105**: 2357–2364. DOI:10.1029/1999JA900442.

Cargill PJ. 2004. On the aerodynamic drag force acting on interplanetary coronal mass ejections. *Sol Phys* **221**: 135–149. DOI:10.1023/B:SOLA.0000033366.10725.a2.

Cargill P, Chen J, Spicer D, Zalesak S. 1995. Geometry of interplanetary magnetic clouds. *Geophy Res Lett* **22**: 647–650.

Cash MD, Biesecker DA, Pizzo V, de Koning CA, Millward G, Arge CN, Henney CJ, Odstrcil D. 2015. Ensemble modeling of the 23 July 2012 coronal mass ejection. *Space Weather* **13**: 611–625. DOI:10.1002/2015SW001232.

Colaninno RC, Vourlidas A, Wu CC. 2013. Quantitative comparison of methods for predicting the arrival of coronal mass ejections at Earth based on multiview imaging. *J Geophys Res : Space Phys* **118**: 6866–6879. DOI:10.1002/2013JA019205.

Coplan MA, Ipavich F, King J, Ogilvie KW, Roberts DA, Lazarus AJ. 2001. Correlation of solar wind parameters between SOHO and

Wind. *J Geophys Res: Space Phys* **106**: 18615–18624. DOI:10.1029/2000JA000459.

Daglis IA. 2001. Space storms, ring current and space-atmosphere coupling in space storms and space weather hazards. In: Proceedings of the NATO Advanced Study Institute on Space Storms and Space Weather Hazards, held in Hersonissos 19–29 June, 2000. Edited by Daglis IA, Crete, Greece: Kluwer Academic Publishers.

Davies JA, Harrison RA, Perry CH, Möstl C, Lugaz N, et al. 2012. A self-similar expansion model for use in solar wind transient propagation studies. *Astrophys J* **750**: 23. DOI:10.1088/0004-637/750/1/23.

Domingo V, Fleck B, Poland AI. 1995. The SOHO mission: an overview. *Sol Phys* **162**: 1–37. DOI:10.1007/BF00733425.

Ebert RW, McComas DJ, Elliott HA, Forsyth RJ, Gosling JT. 2009. Bulk properties of the slow and fast solar wind and interplanetary coronal mass ejections measured by Ulysses: three polar orbits of observations. *J Geophys Res: Space Phys* **114**: A01109. DOI:10.1029/2008JA013631.

Falkenberg TV, Vennerstrom S, Taktakishvili A, Pulkkinen A, Brain DA, Delory GT, Mitchell D. 2010a. CMEs at Earth and Mars. AGU Fall Meeting Abstract.

Falkenberg TV, Vršnak B, Taktakishvili A, Odstrcil D, MacNeice P, Hesse M. 2010b. Investigations of the sensitivity of a coronal mass ejection model (ENLIL) to solar input parameters. *Space Weather* **8**: S06004. DOI:10.1029/2009SW000555.

Gopalswamy N, Lara A, Lepping RP, Kaiser ML, Berdichevsky D, St. Cyr OC. 2000. Interplanetary acceleration of coronal mass ejections. *Geophys Res Lett* **27**: 145–148. DOI:10.1029/1999GL003639.

Gopalswamy N, Lara A, Yashiro S, Kaiser ML, Howard RA. 2001. Predicting the 1-AU arrival times of coronal mass ejections. *J Geophys Res: Space Phys* **106**: 29207–29217. DOI:10.1029/2001JA000177.

Hess P, Zhang J. 2015. Predicting CME ejecta and sheath front arrival at L1 with a data-constrained physical model. *Astrophys J* **812**: 144. DOI:10.1088/0004-637X/812/2/144.

Howard TA, Tappin J. 2009. Reconstructing the 3-D structure and trajectory of ICMEs: physical and forecasting implications. AGU Fall Meeting Abstracts.

Ipavich FM, Galvin AB, Lasley SE, Paquette JA, Hefti S, et al. 1998. Solar wind measurements with SOHO: the CELIAS/MTOF proton monitor. *J Geophys Res* **103**: 17205–17214. DOI:10.1029/97JA02770.

Isavnin A. 2016. FRiED: a novel three-dimensional model of coronal mass ejections. *Astrophys J* **833**: 267. DOI:10.3847/1538-4357/833/2/267.

Kaiser ML, Kucera TA, Davila JM, St. Cyr OC, Guhathakurta M, Christian E. 2008. The STEREO mission: an introduction. *Space Sci Rev* **136**: 5–16. DOI:10.1007/s11214-007-9277-0.

Kleimann J. 2012. 4π odels of CMEs and ICMEs (Invited review). *Sol Phys* **281**: 353–367. DOI:10.1007/s11207-012-9994-8.

Lugaz N, Vourlidas A, Roussev II. 2009. Deriving the radial distances of wide coronal mass ejections from elongation measurements in the heliosphere - application to CME-CME interaction. *Ann Geophys* **27**: 3479–3488. DOI:10.5194/angeo-27-3479-2009.

Manoharan PK. 2006. Evolution of coronal mass ejections in the inner heliosphere: a study using white-light and scintillation images. *Sol Phys* **235**: 345–368. DOI:10.1007/s11207-006-0100-y.

Mays ML, Taktakishvili A, Pulkkinen A, MacNeice PJ, Rastätter L, et al. 2015. Ensemble modeling of CMEs using the WSA-ENLIL+Cone model. *Sol Phys* **290**: 1775–1814. DOI:10.1007/s11207-015-0692-1.

McComas DJ, Elliott HA, Schwadron NA, Gosling JT, Skoug RM, Goldstein BE. 2003. The three-dimensional solar wind around solar

maximum. *Geophys Res Lett* **30**: 1517. DOI:10.1029/2003GL017136.

Michalek G, Gopalswamy N, Chané E. 2002. Arrival time of coronal mass ejections. In Solar Variability: From Core to Outer Frontiers, vol. 506, pp. 177–180.

Millward G, Biesecker D, Pizzo V, de Koning CA. 2013. An operational software tool for the analysis of coronagraph images: determining CME parameters for input into the WSA-Enlil heliospheric model. *Space Weather* **11**: 57–68. DOI:10.1002/swe.20024.

Möstl C, Davies JA. 2013. Speeds and arrival times of solar transients approximated by self-similar expanding circular fronts. *Sol Phys* **285**: 411–423. DOI:10.1007/s11207-012-9978-8.

Möstl C, Rollett T, Lugaz N, Farrugia CJ, Davies JA, et al. 2011. Arrival time calculation for interplanetary coronal mass ejections with circular fronts and application to STEREO observations of the 2009 February 13 eruption. *Astrophys J* **741**: 34. DOI:10.1088/0004-637X/741/1/34.

Möstl C, Amla K, Hall JR, Liewer PC, De Jong EM, et al. 2014. Connecting speeds, directions and arrival times of 22 coronal mass ejections from the Sun to 1 AU. *Astrophys J* **787**: 119. DOI:10.1088/0004-637X/787/2/119.

Odstrcil D, Pizzo VJ. 1999. Distortion of the interplanetary magnetic field by three-dimensional propagation of coronal mass ejections in a structured solar wind. *J Geophys Res* **104**: 28225–28240. DOI:10.1029/1999JA900319.

Odstrcil D, Pizzo VJ, Linker JA, Riley P, Lionello R, Mikic Z. 2004. Initial coupling of coronal and heliospheric numerical magnetohydrodynamic codes. *J Atmos Sol Terr Phys* **66**: 1311–1320. Towards an Integrated Model of the Space Weather System, http://dx.doi.org/10.1016/j.jastp.2004.04.007.

Owens M, Cargill P. 2004. Predictions of the arrival time of coronal mass ejections at 1 AU: an analysis of the causes of errors. *Ann Geophys* **22**: 661–671. DOI:10.5194/angeo-22-661-2004.

Owens MJ, Arge C, Spence HE, Pembroke A. 2005. An event-based approach to validating solar wind speed predictions: high-speed enhancements in the Wang-Sheeley-Arge model. *J Geophys Res: Space Phys* **110**.

Parsons A, Biesecker D, Odstrcil D, Millward G, Hill S, Pizzo V. 2011. Wang-Sheeley-ArgeEnlil cone model transitions to operations. *Space Weather* **9**. DOI:10.1029/2011SW000663.

Pizzo VJ, de Koning C, Cash M, Millward G, Biesecker DA, Puga L, Codrescu M, Odstrcil D. 2015. Theoretical basis for operational ensemble forecasting of coronal mass ejections. *Space Weather* **13**: 676–697. DOI:10.1002/2015SW001221.

Robbins S, Henney CJ, Harvey JW. 2006. Solar wind forecasting with coronal holes. *Sol Phys* **233**: 265–276. DOI:10.1007/s11207-006-0064-y.

Robbrecht E, Berghmans D. 2004. Automated recognition of coronal mass ejections (CMEs) in near-real-time data. *Astron Astrophys* **425**: 1097–1106. DOI:10.1051/0004-6361:20041302.

Rollett T, Möstl C, Isavnin A, Davies JA, Kubicka M, Amerstorfer UV, Harrison RA. 2016. ElEvoHI: a novel CME prediction tool for heliospheric imaging combining an elliptical front with drag-based model fitting. *Astrophys J* **824**: 131. DOI:10.3847/0004-637X/824/2/131.

Rouillard AP, Davies JA, Forsyth RJ, Rees A, Davis CJ, et al. 2008. First imaging of corotating interaction regions using the STEREO spacecraft. *Geophys Res Lett* **35**: L10110. DOI:10.1029/2008GL033767.

Schrijver CJ, Siscoe GL. 2010. Heliophysics: space storms and radiation: causes and effects, Cambridge University Press, Cambridge, UK.

Schwenn R. 1983. The average solar wind in the inner heliosphere: structures and slow variations. In: NASA Conference Publication, vol. 228 of NASA Conference Publication.

Schwenn R. 2006. Space weather: the solar perspective. *Living Rev Sol Phys* **3**: 2. DOI:10.12942/lrsp-2006-2.

Schwenn R, dal Lago A, Huttunen E, Gonzalez WD. 2005. The association of coronal mass ejections with their effects near the Earth. *Ann Geophys* **23**: 1033–1059. DOI:10.5194/angeo-23-1033-2005.

Sheeley NR, Walters JH, Wang Y-M, Howard RA. 1999. Continuous tracking of coronal outflows: Two kinds of coronal mass ejections. *J Geophys Res* **104**: 24739–24768. DOI:10.1029/1999JA900308.

Shi T, Wang Y, Wan L, Cheng X, Ding M, Zhang J. 2015. Predicting the arrival time of coronal mass ejections with the graduated cylindrical shell and drag force model. *Astrophys J* **691**: 806–271. DOI:10.1088/0004-637X/806/2/271.

Stone EC, Frandsen AM, Mewaldt RA, Christian ER, Margolies D, Ormes JF, Snow F. 1998. The advanced composition explorer. *Space Sci Rev* **86**: 1–22. DOI:10.1023/A:1005082526237.

Taktakishvili A, Kuznetsova M, MacNeice P, Hesse M, Rastätter L, Pulkkinen A, Chulaki A, Odstrcil D. 2009. Validation of the coronal mass ejection predictions at the Earth orbit estimated by ENLIL heliosphere cone model. *Space Weather* **7**: S03004. DOI:10.1029/2008SW000448.

Thernisien A, Howard R, Vourlidas A. 2006. Modeling of flux rope coronal mass ejections. *Astrophys J* **652**: 763.

Thernisien A, Vourlidas A, Howard R. 2009. Forward modeling of coronal mass ejections using STEREO/SECCHI data. *Sol Phys* **256**: 111–130.

Vršnak B, Gopalswamy N. 2002. Influence of the aerodynamic drag on the motion of interplanetary ejecta. *J Geophys Res: Space Phys* **107**.

Vršnak B, Temmer M, Veronig AM. 2007. Coronal holes and solar wind high-speed streams: I. Forecasting the solar wind parameters. *Sol Phys* **240**: 315–330. DOI:10.1007/s11207-007-0285-8.

Vršnak B, Žic T, Vrbanec D, Temmer M, Rollett T, et al. 2013. Propagation of interplanetary coronal mass ejections: the drag-based model. *Sol Phys* **285**: 295–315. DOI:10.1007/s11207-012-0035-4.

Vršnak B, Temmer M, Žic T, Taktakishvili A, Dumbović M, Möstl C, Veronig AM, Mays ML, Odstrčil D. 2014. Heliospheric propagation of coronal mass ejections: comparison of numerical WSA-ENLIL+Cone model and analytical drag-based model. *Astrophys J Suppl Ser* **213**: 21. DOI:10.1088/0067-0049/213/2/21.

Yurchyshyn V, Yashiro S, Abramenko V, Wang H, Gopalswamy N. 2005. Statistical distributions of speeds of coronal mass ejections. *Astrophys J* **619**: 599, http://stacks.iop.org/0004-637X/619/i=1/a=599.

Žic T, Vršnak B, Temmer M. 2015. Heliospheric propagation of coronalmass ejections: drag-based model fitting. *Astrophys J Suppl Ser* **218**: 32. DOI:10.1088/0067-0049/218/2/32.

Zhao X, Plunkett S, Liu W. 2002. Determination of geometrical and kinematical properties of halo coronal mass ejections using the cone model. *J Geophys Res: Space Phys* **107**.