Space Weather®

AGU
ADVANCING EARTH AND SPACE SCIENCE

Parameter Distributions for the Drag-Based Modeling of CME Propagation

Gianluca Napoletano[1] , Raffaello Foldes[2,3] , Enrico Camporeale[4,5] ,
Giancarlo de Gasperis[1,6] , Luca Giovannelli[1] , Evangelos Paouris[7,8] ,
Ermanno Pietropaolo[2] , Jannis Teunissen[9] , Ajay Kumar Tiwari[9] , and Dario Del Moro[1]

[1]Dipartimento di Fisica, Università degli studi di Roma "Tor Vergata", Rome, Italy, [2]Dipartimento di Scienze Fisiche e Chimiche, Università degli studi dell'Aquila, Coppito-L'Aquila, Italy, [3]Laboratoire de Mécanique des Fluides et d'Acoustique, CNRS, École Centrale de Lyon, Université Claude Bernard Lyon 1, INSA de Lyon, Écully, France, [4]CIRES, University of Colorado, Boulder, CO, USA, [5]NOAA Space Weather Prediction Center, Boulder, CO, USA, [6]INFN sezione di Roma, Rome, Italy, [7]Institute for Astronomy, Astrophysics, Space Applications & Remote Sensing of the National Observatory of Athens, Penteli, Greece, [8]Faculty of Physics, National and Kapodistrian University of Athens, Athens, Greece, [9]Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands

Abstract In recent years, ensemble modeling has been widely employed in space weather to estimate uncertainties in forecasts. We here focus on the ensemble modeling of Coronal Mass Ejections (CME) arrival times and arrival velocities using a drag-based model, which is well-suited for this purpose due to its simplicity and low computational cost. Although ensemble techniques have previously been applied to the drag-based model, it is still not clear how to best determine distributions for its input parameters, namely the drag parameter and the solar wind speed. The aim of this work is to evaluate statistical distributions for these model parameters starting from a list of past CME-ICME events. We employ LASCO coronagraph observations to measure initial CME position and speed, and in situ data to associate them with an arrival date and arrival speed. For each event we ran a statistical procedure to invert the model equations, producing parameters distributions as output. Our results indicate that the distributions employed in previous works were appropriately selected, even though they were based on restricted samples and heuristic considerations. On the other hand, possible refinements to the current method are also identified, such as the dependence of the drag parameter distribution on the CME being accelerated or decelerated by the solar wind, which deserve further investigation.

Plain Language Summary Coronal Mass Ejections (CME), consisting of huge expulsions of plasma and magnetic field from the solar corona, are important for space weather. Among several forecasting techniques, the drag-based model, which describes CME propagation in interplanetary space, is widely used to compute CME transit time and impact speed, by describing the CME propagation as that of a solid body moving in an external fluid. In recent years, this model has been improved via a new approach in which statistical distributions of the input quantities are introduced to evaluate uncertainties of the resulting forecasts. Unfortunately, such distributions for the model parameters are still not very well known from experimental observations and it is hard to obtain them from theoretical models. In this work, we built an empirical method to evaluate such statistical distributions using a list of past CME-ICME events. New findings emerged from this analysis, such as a dependence of the drag parameter on the interplanetary coronal mass ejections being accelerated or decelerated, deserve further investigation.

1. Introduction

Coronal Mass Ejections (CMEs- P. Chen, 2011; J. Chen, 2017; Webb & Howard, 2012) are expulsions of plasma and magnetic field from the Sun. Their interplanetary counterparts, ICMEs (Kilpua et al., 2017) are among the main drivers of the space weather (SWx-Pulkkinen, 2007; Schwenn, 2006) with impact on the whole heliosphere, and they are responsible for the strongest variations in the near-Earth solar wind conditions (e.g., Buzulukova, 2017; Schwenn et al., 2005; Tsurutani et al., 1988). These variations trigger a number of effects on space-borne and ground-based technologies, either directly or via major geomagnetic storms.

The accurate prediction of the arrival and characteristics of ICMEs at Earth, and more recently elsewhere in the heliosphere, is a necessity to minimize the impact on the existing and future assets, and has always been a

Validation: Raffaello Foldes, Giancarlo Gasperis, Ajay Kumar Tiwari
Writing – original draft: Gianluca Napoletano, Dario Del Moro
Writing – review & editing: Gianluca Napoletano, Raffaello Foldes, Giancarlo Gasperis, Jannis Teunissen, Ajay Kumar Tiwari, Dario Del Moro

primary goal of the SWx forecasting (see e.g., Berrilli et al., 2017; Daglis, 2001; Iwai et al., 2019; Schrijver & Siscoe, 2010; Veettil et al., 2019). Note that in this work the terms CME and ICME refer to the plasma and magnetic field structure expelled from the Sun, without the shock that precedes it.

Forecasting the Time of Arrival (ToA) and Speed of Arrival (SoA) of an ICME more than an hour ahead is nevertheless a complicated task, since it requires to understand the propagation of a poorly determined plasma and magnetic field structure into an essentially undetermined interplanetary environment. A recent review (Vourlidas et al., 2019) assessed the current state of ICME ToA and SoA forecasting algorithms by surveying the recent literature. While there is quite a scatter in the results and perhaps a bias correlated to the sample size used, the authors report that the most recent forecasting methods have a mean absolute error (MAE) close to 10 hr, which is similar in the case of empirical, simplified physics or full MHD models (as ENLIL or EUHFORIA Odstrcil, 2003; Pomoell & Poedts, 2018), or machine learning based (see Camporeale, 2019, for a list of such approaches). Vourlidas et al. (2019) conclude that a number of factors (physical, observational and modeling) are limiting the performances of all approaches, and that the ToA accuracy in particular is limited by the quality of the currently available data. These complications arise from the difficulty to evaluate remotely the properties of the CME at launch with the present-day instrumentation and from the impossibility to properly characterize the status of the inner heliosphere. Therefore, in order for the forecast to be useful, it should cope with this lack of information and provide an estimate of its intrinsic uncertainty (Owens et al., 2020). This can be achieved by empirical methods through the use of statistical relationships established between past CME measured parameters and ICME characteristics (Gopalswamy et al., 2001; Kilpua et al., 2012), or can be achieved by numerical MHD-based models by using ensembles of runs to model the same ICME with different initial conditions that represent the inherent uncertainties (Cash et al., 2015; Emmons et al., 2013; Mays et al., 2015).

The main shortcoming of the statistical approach is that it "treats all events the same and neglects the contextual information and knowledge that certain situations are inherently more predictable than others" (Owens et al., 2020). The other approach (i.e., treat every CME as a different case) has to cope with a relatively large parameter space to explore, and the relatively long time needed for the computation of each simulation run (of the order of tens of minutes on high-performance systems).

A possible solution is to adopt simplified, kinematic models, for instance, assuming a simplified solar wind propagation, a simple CME geometry, and a hydrodynamic-like ICME-solar wind interaction. Among these models, the drag-based model (DBM-Cargill, 2004; Vršnak et al., 2013), is among the most used and it can be run in large ensembles during a few seconds in an average laptop (Amerstorfer et al., 2018; Dumbović et al., 2018; Kay & Gopalswamy, 2018; Napoletano et al., 2018). The DBM requires CME properties as input: the CME launch time, initial speed, direction and angular width.

To model the interaction of the ICME with the background solar wind the DBM needs only the solar wind speed and the value of the drag parameter γ, which determines the interaction between the solar wind and the CME. The quantities used to describe the CME are retrieved from observations that have associated measurement errors. They can be used in ensemble models assuming Gaussian probability distribution functions (PDFs). The solar wind speed and the drag parameter are instead drawn from *a-priori* PDFs, modeled from empirical PDFs built from past data sets of CME and associated ICME characteristics measured at liftoff and at Lagrange point L1, respectively. The outputs of the DBM ensemble model are the PDFs of ToA and SoA at a target location. From these, we can estimate the most probable ToA and SoA, and their associated prediction uncertainties (e.g., Del Moro et al., 2019; Piersanti et al., 2020).

In recent years there has been several interesting development of DBM based tools to forecast ICME arrival time and speed. When compared against other forecasting methods as in (Vourlidas et al., 2019), their performance is comparable and sometimes even better than MHD based methods (Vršnak et al., 2014). Augmenting the results presented in Table 1 of (Vourlidas et al., 2019) with Table 2 of (Dumbovic et al., 2021), we can estimate that the typical MAE of the DBM methods is of the order of 10 hr and the typical error on the SoA is around 50 km/s. It deserves to be stressed that the performance of each model has been computed on a different ICME data set, thus limiting the value of this comparison. A community effort to create a common benchmark is underway (e.g., Verbeke et al., 2019), but it is not yet a widespread standard.

Recently, Kay et al. (2020) used the DBM ensemble approach to answer an implicit question in Vourlidas et al. (2019): "How much should we improve our knowledge of the parameters to improve the ToA predictions

beyond the present limit MAE of about 10 hr"? Their conclusion is that the most critical parameter is the CME speed at liftoff, and efforts should be spent to achieve a better and more homogeneous determination of the CME speed via coronagraphics imaging or other means. Kay et al. (2020) also explored the sensitivity of the ToA versus the γ parameter, following Dumbović et al. (2018), who postulated a symmetric distribution of values ($\gamma = 0.1 \pm 0.05 \times 10^{-7}$ km^{-1}). However, Napoletano et al. (2018) (hereafter Paper I) built an empirical γ PDF which is asymmetrical and not compatible with a Gaussian shape. Since the γ parameter incorporates much of the physics of the ICME-wind interaction and its precise value is poorly understood (Kay et al., 2020), we think it deserves further investigation. In particular, the empirical γ PDF in Paper I has been built from a limited database of ICME, therefore it is the quantity that most of all needs a more robust definition. A better assessment of the values of γ may be obtained, in principle, with detailed knowledge of the ICME kinematics. In order to evaluate the DBM input parameters, (Žic et al., 2015) proposed to employ a least squares fitting procedure to the ICME interplanetary tracking derived from STEREO coronagraphic and heliospheric image observations. In addition, (Rollett et al., 2016) also developed such an approach and employed an improved model for the geometrical shape of the CME front, showing that the extrapolation of the CME dynamics based on real-time tracking can further reduce the mean error of the predicted arrival time to 6.4 $\pm$ 5.3 hr and impact speed to 16 $\pm$ 53 km/s. Although promising, such an approach is currently not feasible for real-time forecasting, as it requires a dedicated heliospheric observatory for real-time ICME tracking. Therefore model parameters have to be constrained by two single observations (at liftoff and Earth), and a better assessment of these parameters depends on the availability of more data and an improved database. The aim of this paper is to use a large number of ICMEs to build a new empirical PDF for the model parameters and to find a suitable functional form to model it.

The paper is structured as follows. In Section 2 we describe the data and the methods employed to build the ICME database used to obtain the new γ PDF. Section 3 contains the technical description of the methods used to retrieve the γ PDF. Section 4 contains the actual results and a validation of those results against the ICME list from Paouris and Mavromichalaki (2017a). In Section 5 we discuss the results and provide a synopsis of the findings of this paper. The ICME catalog built for this analysis, together with a tool for the data set visualization and the software employed for the ensemble simulation through the probabilistic drag based model, is available from https://doi.org/10.5281/zenodo.5818470 (Napoletano et al., 2021).

2. Data

As mentioned above, the DBM needs input values to output a ToA and a SoA at a target location. For the purpose of this work, we compare observed ToA and SoA at Earth (or L1) position against those computed by the probabilistic drag-based model (PDBM) presented in Paper I. We therefore need a database associating Earth ToA and SoA of an ICME to the kinematic characteristics of the corresponding CME. In particular, we need measures of the position r_0 and the speed v_0 at time t_0 of the CME front, and the solar wind speed (see A3 in Appendix A).

Also, to obtain a homogeneous database for a consistent approach, we re-computed some of the CME kinematics properties (CME leading edge initial initial speed and acceleration) with a standardized method. Last, to have an assessment of the method as close as possible to the actual operation performances, we made use of the same methods and algorithms presently implemented in the real-time ICME ToA forecast running at http://spaceweather.roma2.infn.it. These selection criteria and methods are described in the following sub-sections.

2.1. Sources for Compiling the Employed ICMEs Database

Our analysis uses a database connecting the kinematic parameters of the CME at launch time and the information about the arrival time and speed of the related ICME. The databases already available are not suited for our analysis, but we can merge and complement the information from three different sources available online to create a new database for this purpose.

We start from the list of near-Earth interplanetary coronal mass ejections (ICMEs) compiled by Richardson and Cane (2010), hereafter R&C, 2010, who maintain the catalog using data from the OMNI database (Goddard Space Flight Center, GSFC, http://spdf.gsfc.nasa.gov/). From about 500 events from May 1996 to present in this catalog, we consider only those events that have information about the CME liftoff time, selecting 247 events. This onset time (see http://www.srl.caltech.edu/ACE/ASC/DATA/level3/icmetable2.htm for specifications) refers

Table 1
Column Description of the ICMEs Data Set Created in This Work

	Name	Keyword	Description	Source
1	LASCO start	LASCO_Start	First CME appearance in LASCO C2/C3 coronographs	LASCO/CDAW
2	Start date	Start_Date	Start time of CME extrapolated at 20 $R_\odot$	This work
3	Arrival date	Arrival_Date	Estimated arrival time of the ICME based primarily on plasma and magnetic field observations	R&C
4	Plasma event dur.	PE_duration	End of the ICME plasma signatures after col. 3 is recorded	R&C
5	Arrival speed	Arrival_v	(km/s) ICME arrival speed measured at L1 (∼1AU)	R&C
6	Transit time	Transit_time	(hrs.) Computed between col. 3 and col. 1	This work
7	Trans. time error	Transit_time_err	(hrs.) Error associated to the extrapolated start date (col. 3) of a CME	This work
8	LASCO date	LASCO_Date	Most likely CME associated with the ICME observed by LASCO	LASCO/CDAW
9	LASCO speed	LASCO_v	(km/s) Max. plane-of-sky (POS) CME speed along the angular width	LASCO/CDAW
10	Position angle	LASCO_pa	(deg.) Counterclockwise (from solar North) angle of appearance into coronographs	LASCO/CDAW
11	Angular width	LASCO_da	(deg.) Angular expansion of CME into coronographs	LASCO/CDAW
12	Halo	LASCO_halo	If col. 15 is >270° then 'FH' (full halo), if >180° 'HH' (half halo), if >90° 'PH' (partial halo), otherwise 'N'.	LASCO/CDAW
13	De-proj. speed	v_r	(km/s) De-projected CME speed (from 9, see Appendix A1)	This work
14	De-proj. speed error	v_r_err	(km/s) Uncertainty of CME initial speed (col. 13)	This work
15	Theta source	Theta	(arcsec) Longitude of the most likely source of CME	This work
16	Phi source	Phi	(arcsec) Co-latitude of the most likely source of CME	This work
17	Source pos. error	POS_source_err	(deg.) Uncertainty of the most likely CME source	This work
18	POS source angle	POS_source	(deg.) Principal angle of the most likely CME source	This work
19	Relative width	rel_wid	(rad.) De-projected width of CME	This work
20	Mass	Mass	(g) Estimated CME Mass (if provided)	LASCO/CDAW
21	Solar wind type	SW_type	Solar wind (slow, S, or fast, F) interacting with the ICME	This work
22	Bz	Bz	(nT) z-component of magnetic field at L1 and CME arrival time (col. 3)	R&C
23	Dst	DST	Geomagnetic Dst index recorded at CME arrival (col. 3)	R&C
24	Stat. de-proj. speed	v_r_stat	(km/s) Statistical de-projected CME speed, that is, v_r_stat = LASCO_v*1.027 + 41.5	(Paouris & Mavromichalaki, 2017b)
25	Acceleration	Accel	(m/s^2) Residual acceleration at last CME observation	This work
26	Analytic sol. wind	Analytic_w	(km/s) solar wind from DBM exact inversion	This work
27	Analytic gamma	Analytic_gamma	(km^{-1}) drag parameter, γ, from DBM exact inversion	This work

to the most-likely association between the ICME and the corresponding CME observed by LASCO coronagraphs (C2 and C3) aboard SOHO spacecraft (in a few cases, STEREO spacecraft observations are used).

We make use also of information about the CME liftoff retrieved from the SOHO LASCO CME catalog (https://cdaw.gsfc.nasa.gov/CME_list/) maintained at the CDAW Data Center. Lastly, information about features and events on the Sun that may be associated to the CME, such as Filaments, Flares, Active Regions, and Coronal Holes are retrieved from the Heliophysics Event Knowledgebase (HEK; Hurlburt et al., 2010; Martens et al., 2012): a queryable repository of feature and event information about the Sun.

2.2. Database Creation

We associate each of the remaining 247 ICME in the R&C list with an entry in the CDAW SOHO/LASCO CME catalog having the same onset times, $t_{0_{R\&C}} = t_{0_{LASCO}}$, by using a dedicated Python routine. The SOHO/LASCO catalog provides various kinematic properties associated to the onset time $t_{0_{LASCO}}$ of the CMEs (i.e., position angle, angular width, Mass, see Table 1). Among those properties, there is also an estimate for the CME lift-off speed, but unfortunately, it is relative to the fastest part of the CME front and has no associated error. Therefore,
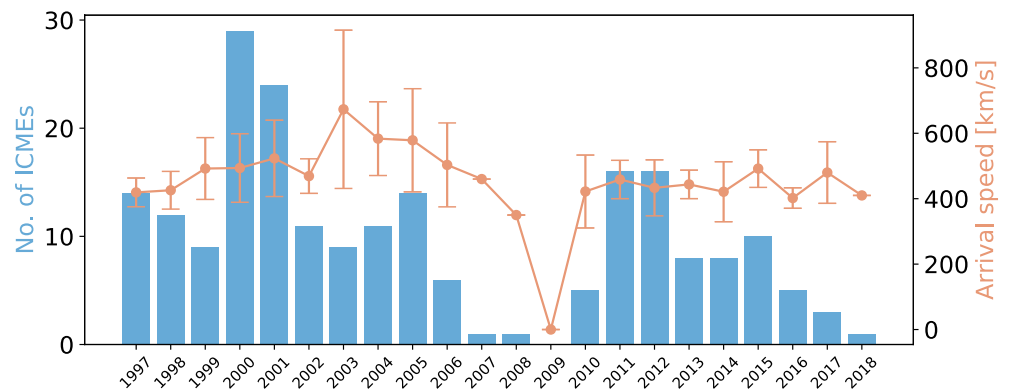
Figure 1. Distribution of the selected interplanetary coronal mass ejections (ICMEs) in our data set during the past 20 years. The orange line represents the average Speed of Arrival of ICMEs for each year, the error bar represents the relative standard deviation.

it can not be employed to define a PDF and is not suitable for our purposes. Nevertheless, with this association, we can retrieve from the SOHO/LASCO website the original height versus time measurements of the CME front (in solar radii units, R_{Sun}), computed from LASCO C2/C3 coronagraph images. These values represent the position of the CME front on the plane-of-sky (hereafter POS), and need to be de-projected to obtain the true radial distance and speed of the CME. The equation to de-project the position $r_{0_{POS}}$ into r_0 is from (Gopalswamy et al., 2010). It implies a model for the CME shape (i.e., the CME front expansion is considered completely radial, as in Model A of Figure 9 in Schwenn et al., 2005) and requires the CME angular width and the location of the CME source on the solar disk.

The details about the algorithms to associate a CME to its source on the solar disk and to compute its de-projected speed at $r_0 = 20R_{Sun}$ and the associated error are described in Appendix A.

An interesting new geometrical technique for de-projecting the CME speed has been very recently published (Paouris, Vourlidas, et al., 2021). We foresee the possibility to utilize such a method in a future work. After these procedures our data set was reduced to 214 CME-ICME pairs. In Figure 1 we report the histogram of the yearly number of CME comprising this data set, since it spans about two solar cycles (23 and 24). As it is known (e.g., Lamy et al., 2019; Webb & Howard, 2012), the number of CMEs depends on the solar cycle phase and the total number of CMEs per cycle is clearly different for the cycle 23 and 24. In Figure 2 we show a summary of some quantities reported in Table 1, and in particular we focus on those variables that will be used for the inversion procedure (see Section 3.2).

3. Methods

3.1. The Drag-Based Model

We employ the drag-based model (DBM; Cargill, 2004; Vršnak et al., 2013) to forecast arrival time and impact speed of a CME at Earth. This model assumes that, from a certain distance from the Sun, the CME dynamics is governed only by its interaction with the ambient solar wind. By employing a fluid dynamic analogy, it is assumed that the force depends on the square of the relative velocity of the CME to the ambient solar wind flow, so that the equation for the CME radial acceleration reads:

$$a = -\gamma(r)\left[v - w(r)\right]\left|v - w(r)\right| \tag{1}$$

where $\gamma(r)$ is the so-called drag parameter, representing the interaction efficiency between the CME and the solar wind, $w(r)$ is the solar wind speed, and r is the distance from the Sun. A reasonable approximation beyond 20 solar radii is that of constant γ and w throughout the whole ICME propagation (Cargill, 2004; Vršnak et al., 2013). We point out that this is a relevant assumption, as in reality the ambient solar wind speed and the mechanisms of interaction between the solar wind and the CME structure are not constant. We refer the reader for example, to (Temmer et al., 2012), (Rollett et al., 2014), (Žic et al., 2015) for studies investigating CMEs evolving in different drag regimes and variable solar wind speed, or (Piersanti et al., 2020) for an example of the probabilistic approach
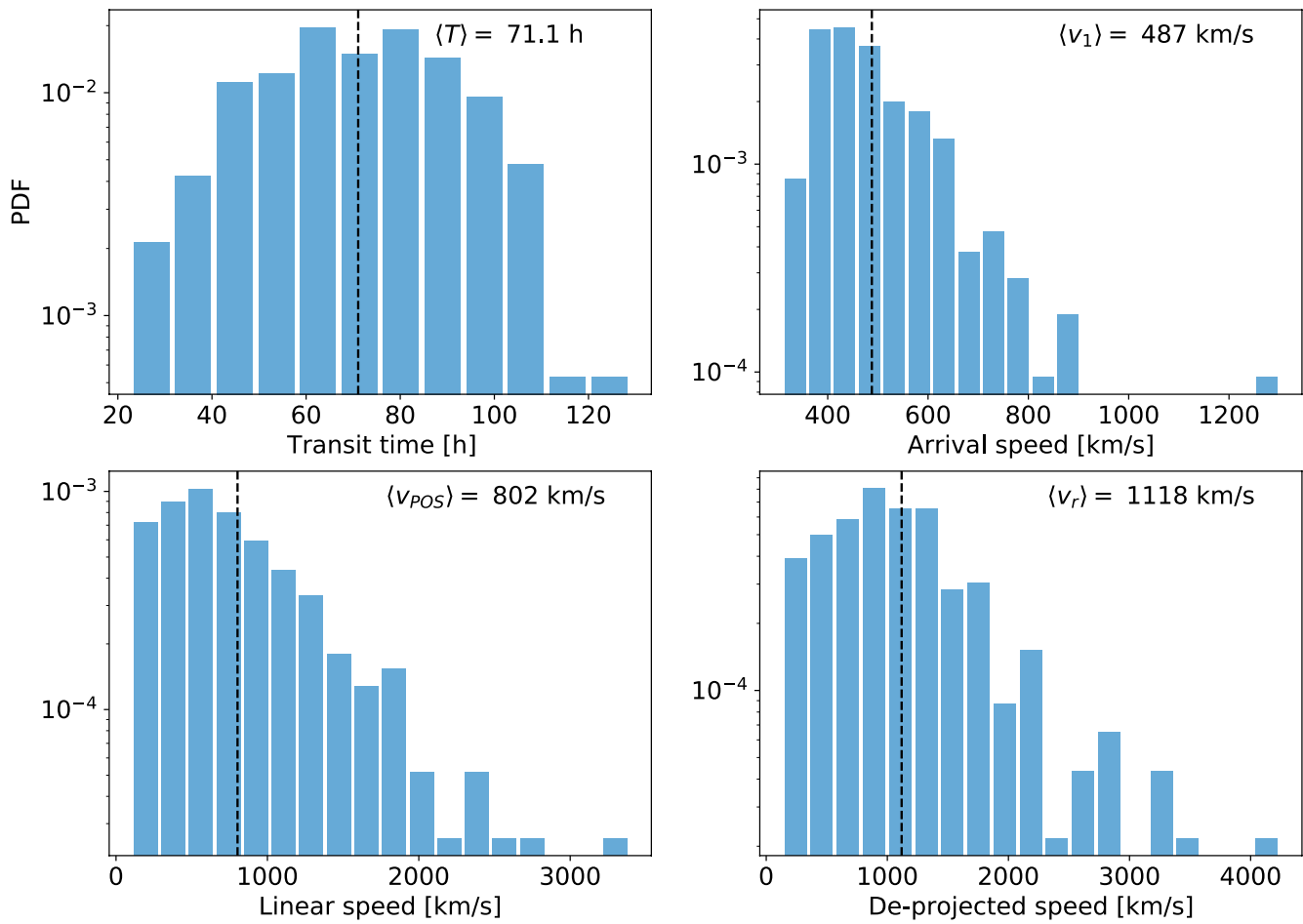
Figure 2. (From left to right) Probability distribution of the interplanetary coronal mass ejections (ICME) Time of Arrival, Speed of Arrival, linear plane-of-sky liftoff speed and de-projected liftoff speed obtained with from whole ICME data set. In every panel, the vertical dashed line (black) is the average value, also reported within each panel.

to the drag-based model applied with a variable solar wind. Under such assumptions, Equation 1 can be solved analytically (Vršnak et al., 2013) for the heliospheric distance and the ICME speed as a functions of time:

$$r(t) = \pm \frac{1}{\gamma} \ln \ [1 \pm \gamma(v_0 - w)t] + wt + r_0, \tag{2}$$

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

where the choice between $\pm$ depends on the sign of $v_0 - w$: the + sign is taken for accelerated CMEs ($v_0 - w < 0$), while the - sign holds for decelerated ones ($v_0 - w > 0$). Given the initial conditions r_0, v_0 and model parameters γ and w, these equations can be employed to compute the ToA and the SoA at a target located at a chosen helio-centric distance.

3.2. The Inversion Procedure

The DBM Equations 2 and 3 can be used to forecast the travel time T and the impact speed v_1 of an ICME at a given position r_1. Conversely, if T and v_1 are known, these equations can be inverted leaving γ and w as unknown values, as in (Vršnak et al., 2013):

$$r_1 = r_0 + wT + \frac{(v_0 - w)(v_1 - w) \ T}{(v_0 - v_1)} \ln \left[1 + \left(\frac{v_0 - v_1}{v_1 - w} \right) \right] \tag{4}$$

$$\pm\gamma = \frac{(v_0 - v_1)}{(v_1 - w)(v_0 - w)T} \quad (5)$$

the first equation is implicit in w, but can be solved numerically and its solution is then employed to compute γ through the second one. Therefore, for a given T, v_0, v_1, we have $\gamma = f(w)$. This dependence can be seen in the joint PDFs presented below, where for each CME there is a ridge of joint w, γ values.

In principle, using the entries of the database presented in Section 2, we could compute the γ and w values for every ICME and add them to the database. In practice, due to the intrinsic errors of the r_0, r_1, v_0, v_1 and T parameters, or due to the fact that sometimes the DBM model simply does not accurately represent the ICME motion (e.g., in the case when $w = $ constant is not a viable approximation) Equations 4 and 5 have no solution for 75 out of the 214 events of our database. We therefore used an inversion method that takes into consideration the experimental uncertainty of the input quantities during the inversion procedure. The two parameters r_0 and r_1 have no associated errors, since we set them at 20 and 215 solar radii, respectively. For v_0, v_1 and T, we modeled the PDFs with normal distributions centered at the relative measured or estimated values and with standard deviations corresponding to their uncertainties. For v_0, the mean is defined by the CME de-projected speed, and the standard deviation σ by its associated error, rows 13 and 14 in Table 1 respectively. For v_1, the mean is defined by the ICME measured arrival speed (row 5 in Table 1), and σ by an assumed measurement error of 10%. For T, the mean is defined by the difference between the measured arrival time at Earth and the estimated passage at $20R_{Sun}$ row 6 in Table 1, and σ by considering the error on the liftoff time obtained by the de-projection procedure (row 7 in Table 1).

From the average values and their PDFs, the inversion method generates N random samples of $[r_0, r_1, v_0, v_1, T]$ per ICME and feeds those to Equations 4 and 5. We run this inversion procedure with $N = 5,000$ for each of the 214 events in the ICME database. In about 50% of all the generated cases the solutions γ and w do not exist due to incompatibility of the randomly generated input values, that is, such values do not allow for a solution of the implicit Equation 4.

Also, we take into consideration for the following analysis only those solutions where $10^{-8}\,\mathrm{km^{-1}} < \gamma < 10^{-6}\,\mathrm{km^{-1}}$. We choose this range of magnitudes considering Equation 2 of Vršnak et al. (2013) and following the same order of magnitude reasoning therein, which poses a limit on the realistic values of γ.

4. Results

4.1. Inversion Method Results

The inversion procedure was successful for 210 out of 214 events, thus providing a statistical distribution for the solar wind w and drag parameter γ. The whole sample consists of 519,857 inversions. In the upper-left panel of Figure 3, we show the joint distribution $\gamma - w$. As already stated in Par. 3.2, in the joint distribution we can still identify a few w–γ ridges generated by single CMEs, but the plane is populated enough to extract the properties of these PDFs and compare them with the PDFs used in Paper I. Also, as a consequence of the random extraction of the initial speed and solar wind, we can draw two more joint PDFs, separating the accelerated $\Delta v = v_0 - w < 0$ (18,416 extractions) and decelerated $\Delta v > 0$ (501,441 extractions) ICMEs. These two joint PDFs $\gamma - \Delta v$ are shown respectively in the central-left and lower-left panels of Figure 3.

### 4.2. PDFs for the Solar Wind Speed and Drag Parameter

From the joint distributions shown in Figure 3, we can extract the marginal distributions for the drag-parameter, γ. We compare this empirical PDF with the reference lognormal function used in Paper I to model the γ PDF. The plot in the upper panel of Figure 3 shows that the histogram retrieved from the whole data set has a shape which is not compatible with the reference PDF (in blue). The fit of the histogram with a lognormal function (red dashed) retrieves $\mu = -0.83$ and $\sigma = 1.26$, against the $\mu = -0.70$ and $\sigma = 1.01$ of the reference PDF. It is worth to note that we tried fitting to the PDFs other function types, in particular exponentials, power laws, and truncated power laws, but none of those retrieved a better fit than the lognormal.

As above, we can divide our data set in accelerated and decelerated ICMEs (central left and lower left panels of Figure 3). The two histograms look quite different and are fitted by lognormal functions (red dashed) with
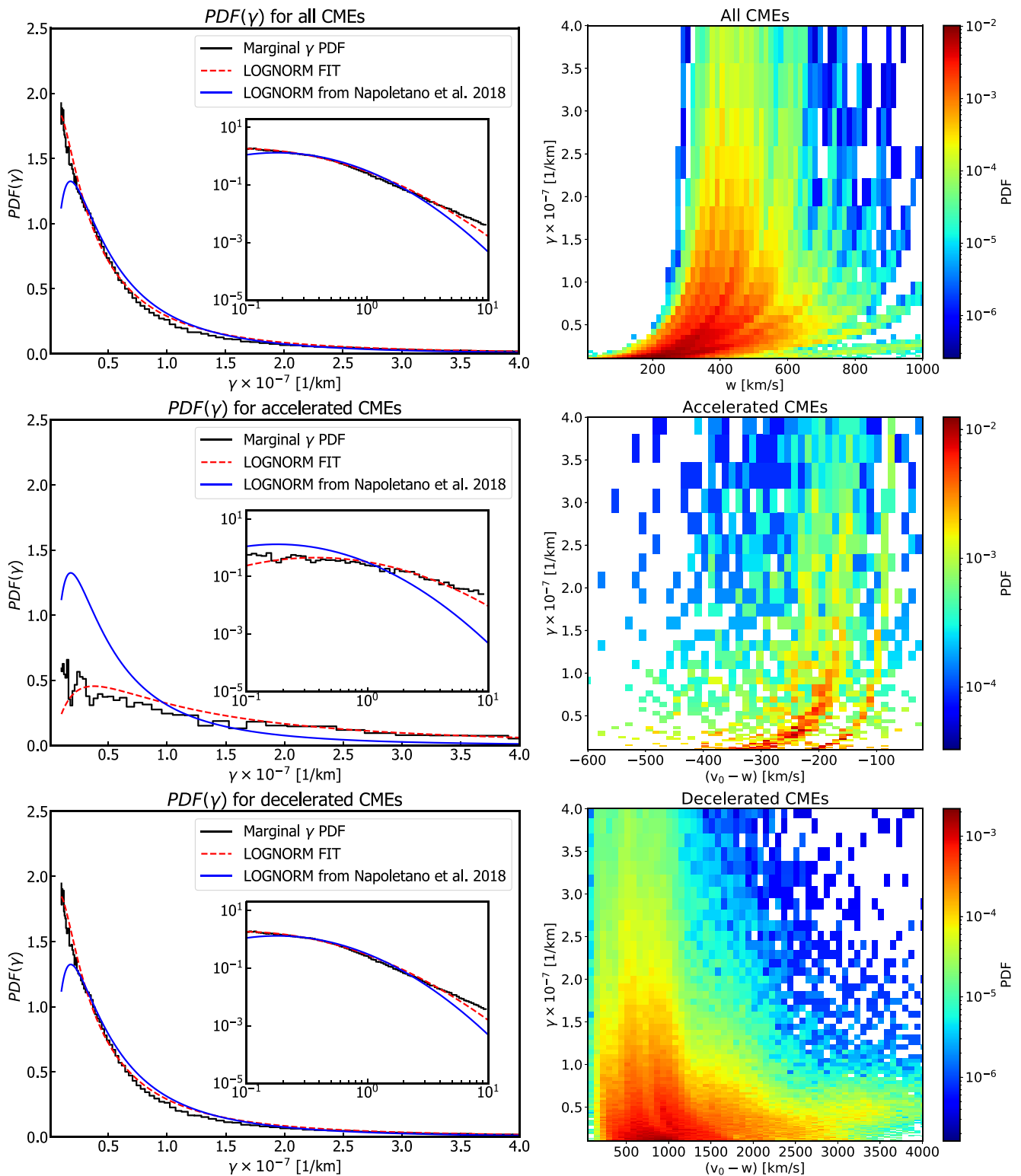
Figure 3. Probability distribution functions (PDFs) for γ (left column) and the joint distributions (right column) obtained from inversion procedure using the interplanetary coronal mass ejections (ICMEs) in the database as input. Upper panel: γ versus w for the whole data set (519,857 values). Middle panel: γ versus Δv for the accelerated ICMEs (18,416 values). Lower panel: γ versus Δv for the decelerated ICMEs (501,441 values). All the PDFs were fitted with a lognormal function (red dashed). The lognormal function (blue line) used in Paper I is also plotted for comparison. In each panel, the inset shows the same plot on a log-log scale for an extended range.
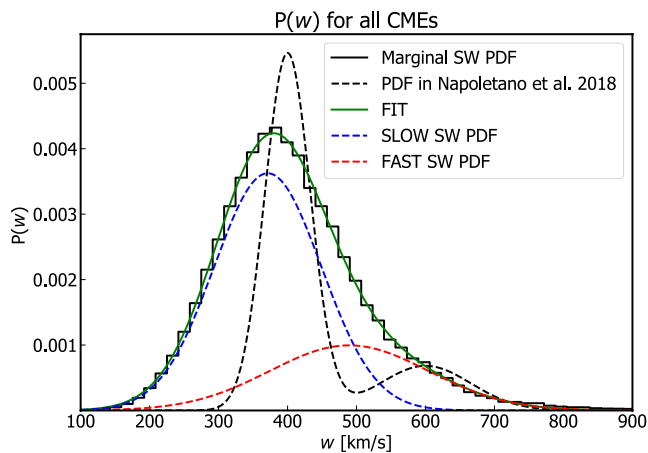
Figure 4. The probability distribution function (PDF) of the solar wind speed from the whole data set. The PDF was fitted with the sum of two Gaussian functions (green line), which can be interpreted as the slow solar wind (blue dashed line) and fast solar wind (red dashed line) contribution to the whole PDF(w). As a reference, we also plot the SW PDF used in Paper I (black dashed line).

significantly different parameters. In particular, from the fit to the histogram from the decelerated ICMEs, we retrieve $\mu = -0.85$ and $\sigma = 1.25$. Since this histogram contains more than 98% of the total samples, its parameters are close to those obtained from the whole data set, but also more compatible with the values of Paper I.

From the fit to the histogram from the accelerated ICMEs, we retrieve $\mu = 0.40$ and $\sigma = 1.18$. The accelerated CMEs are represented by a lognormal distribution with a significantly higher mean value.

From these results, we define new γ parameter distributions functions for accelerated and decelerated cases. Similarly, we can extract the partial distribution of the solar wind values obtained by all the inversion procedures black line in Figure 4. We fitted this distribution with the sum of two Gaussian functions (blue and red lines), whose parameters are reported in Table 2. It is straightforward to interpret these two Gaussian functions in terms of slow and fast solar wind. We therefore redefine the parameters to generate the PDFs for the solar wind employed in Paper I with the new values reported in Table 2.

4.3. Testing PDBM Performance With Updated Gamma and Solar Wind Speed Distributions

In order to test the new solar wind and drag parameter distributions and compare the performance of the PDBM related to old distributions (Paper I) our ICMEs list was used to compare the predictions with the observed values. The algorithm maps a large sample of initial conditions to the corresponding transit time and arrival speed through the DBM Equations 2 and 3, allowing, in addition, to asses the forecast uncertainty from the output distributions. For each event, we start by randomly extracting an initial speed v_0 and a value for the solar wind speed w from their respective normal distributions. If $v_0 < w$ ($v_0 > w$) the event is an accelerating (decelerating) CME, and a value for the drag parameter γ is randomly generated from the corresponding log-normal distribution from Section 4.2. Computed CME transit time T_c and impact speed at 1AU v_C are then obtained through the DBM equations with this set of initial conditions and parameters. For each event, a statistical distribution for ToA and SoA results from repeating this procedure a large number of times. The mean values $\langle T_c \rangle$ and $\langle v_c \rangle$ of such distributions, together with their standards deviations are taken as representative of the forecast for each event.
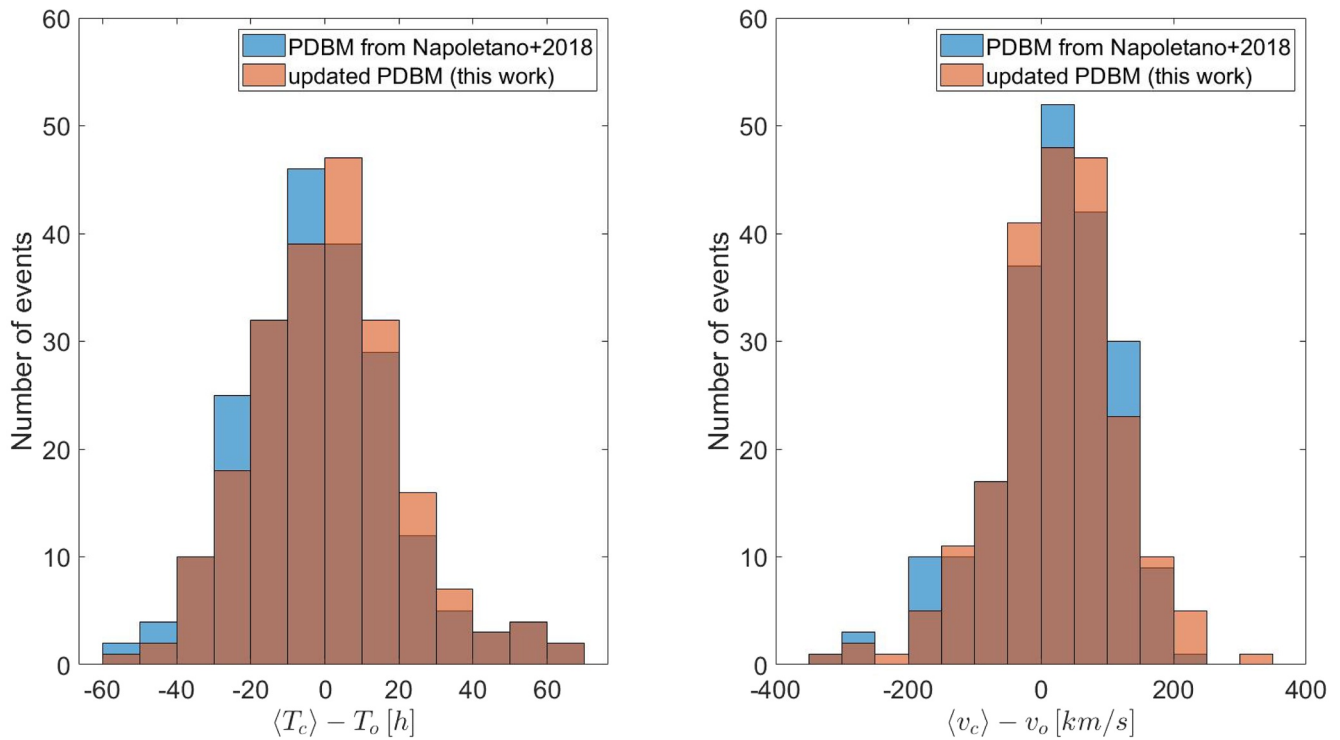
Figure 5 shows the histograms of the difference between transit time and arrival speed comparing the old PDBM and the new one on this set of events. Relevant performance indicators are collected in Table 3. Figure 6 shows the plots of the PDBM prediction versus the observed values of the ICME transit time and impact speed. For 74% of events, the observed ToA falls within the standard deviation of the predicted one, and for the arrival speed this occurs for 90% of the events. This may be an indication of a possible overestimation of the error on the forecasted ToA, related to the errors on the input velocities.

4.4. Validation Against a Subset of Paouris' ICME Database

To test the validity of the new PDBM distributions on a list of past ICME events which is independent from the one described in Section 2, we employed a list of 100 events from the (Paouris & Mavromichalaki, 2017a) list, obtained after excluding 92 common ICMEs between the two databases. Figure 7 shows the histograms of the difference between transit time and arrival speed comparing the old PDBM and the new one on this set of events. Relevant performance indicators are collected in Table 4. Figure 8 shows the plots of the PDBM prediction versus the observed values of the ICME transit

Table 2
Parameters for the Two Gaussian Functions Used to Model the Solar Wind Distribution

	A	$\overline{w}$ [km/s]	σ_w [km/s]	
Blue Gaussian function	$3.6 \cdot 10^{-3}$	370	80	slow solar wind
Red Gaussian function	$1.0 \cdot 10^{-3}$	490	100	fast solar wind

Figure 5. Histograms of the differences between computed and observed values for the interplanetary coronal mass ejections (ICME) travel time (left) and arrival speed (right), computed using our ICME list. Such predicted values are computed both with the old solar wind and drag parameter distributions from Paper I and the new ones.

time and impact speed. For 65% of events, the observed ToA falls within the standard deviation of the predicted one, while for the arrival speed this occurs for 78% of the events.

5. Discussion and Conclusions

We compiled a list of CME-ICME pairs with a reliable association between their remote observations and in situ signatures. To evaluate the CME initial speed and launch time, we applied a polynomial fit procedure to coronagraph data. We then employed an algorithm to establish the most suitable type of solar wind accompanying the CME. We built this database by using the automated methods for CME characterization that are used for a CME detection and forecast service. Therefore, the uncertainties reported in the database should be representative of the errors in real-time use.

We employed this database to retrieve the parameters γ and w in the DBM equations. In this procedure, we took into account the uncertainty on the observations, mapping the input PDFs into the PDFs for the model

Table 3
Performance Indicators for the Application of the PDBM Model From Paper I and With the Updated PDFs to the Database Presented in This Work

		PDBM from paper I	New PDBM
ME of arrival time prediction	$\langle \Delta T \rangle$ [h]	-1.7 ± 21.3	1.1 ± 20.6
MAE of arrival time prediction	$\langle |\Delta T| \rangle$ [h]	16.4	15.9
RMSE of arrival time prediction	RMSE(T)[h]	21.3	20.4
ME of arrival speed prediction	$\langle \Delta v_1 \rangle$ [km/s]	14.2 ± 105.7	20.3 ± 103.7
MAE of arrival speed prediction	$\langle |\Delta v_1| \rangle$ [km/s]	79.1	78.0
RMSE of arrival speed prediction	RMSE(v_1)[km/s]	106.4	111.7
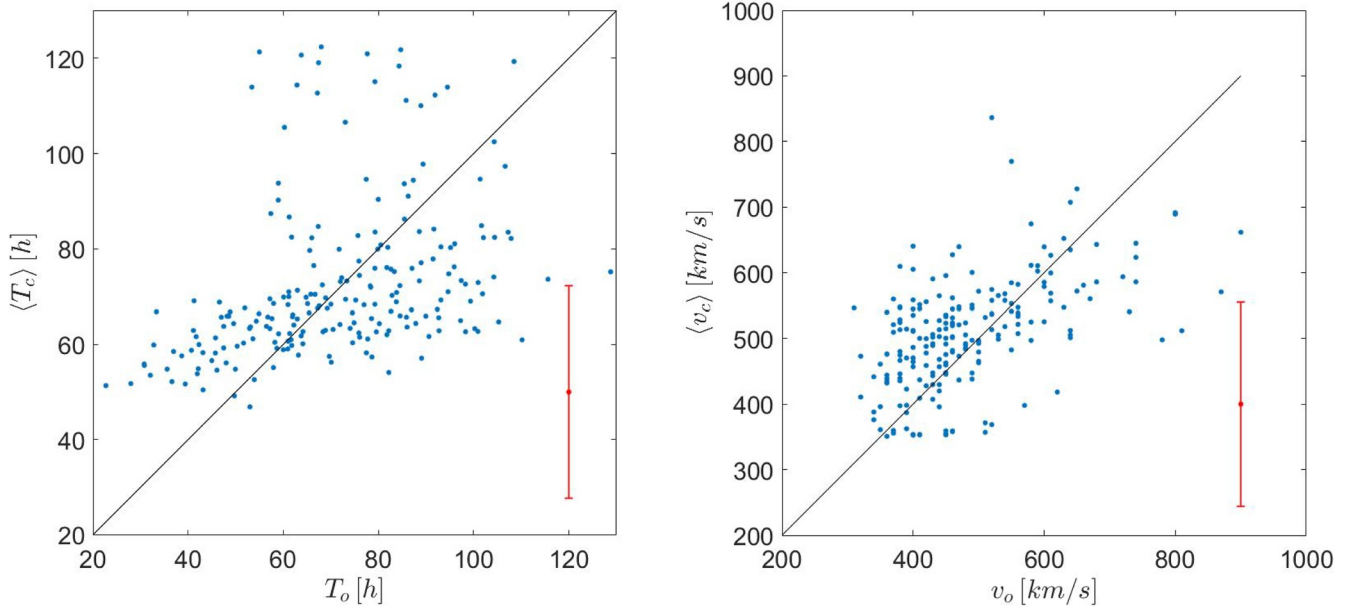
Figure 6. Plots of the computed mean Time of Arrival (ToA) versus the observed ToA (left) and of the computed mean Speed of Arrival (SoA) versus the observed SoA (right), obtained after applying the new PDBM distributions on our data set. Black line represents the perfect forecast. Due to the density of data points, vertical error bars have been omitted for sake of visualization, and the median error bar has been represented in the bottom right corner.
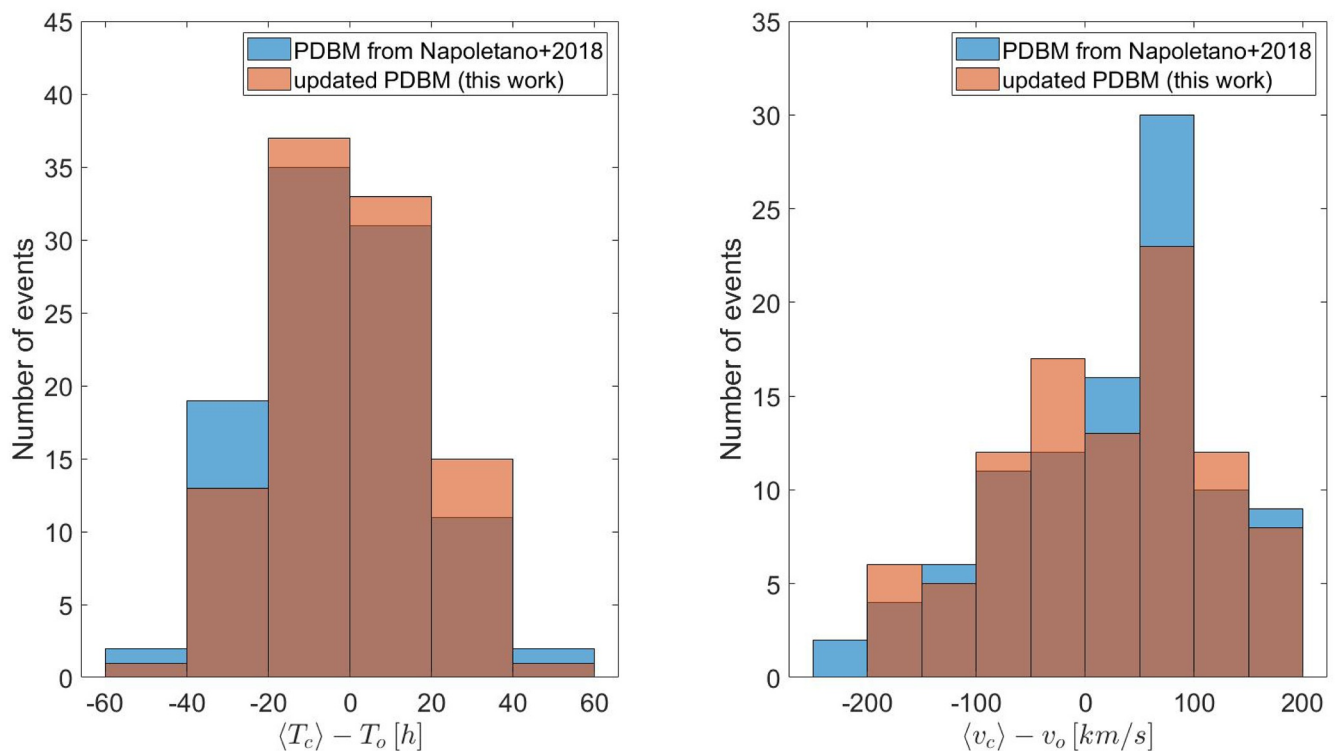
Figure 7. Histograms of the differences between computed and observed values for the interplanetary coronal mass ejection travel time (left) and arrival speed (right), computed on on the subset of events from (Paouris & Mavromichalaki, 2017a) list. Such predicted values are computed both with the old solar wind and drag parameter distributions from Paper I and the new ones.

Table 4
Performance Indicators for the Application of the PDBM Model From Paper I and With the Updated Probability Distribution Functions to the Database by (Paouris & Mavromichalaki, 2017a)

		PDBM from paper I	New PDBM
ME of arrival time prediction	$\langle \Delta T \rangle$ [h]	-3.0 ± 19.9	-0.2 ± 19.5
MAE of arrival time prediction	$\langle |\Delta Tl \rangle$ [h]	16.8	16.3
RMSE of arrival time prediction	RMSE(T) [h]	14.4	15.5
ME of arrival speed prediction	$\langle \Delta v_1 \rangle$ [km/s]	24.4 ± 97.0	30.0 ± 102.4
MAE of arrival speed prediction	$\langle |\Delta v_1| \rangle$ [km/s]	84.1	88.8
RMSE of arrival speed prediction	RMSE(v_1)[km/s]	71.9	80.4

parameters. The robust statistics, granted both by the larger ICME database with altogether 214 ICMEs and the Monte-Carlo-like inversion method, produced a joint PDF populated enough to allow us to verify the PDFs proposed in (Napoletano et al., 2018).

The empirical solar wind w PDF has been modeled using the sum of two Gaussian functions, and given their parameters we interpreted these as representative of the slow and fast solar wind distributions.

Similarly, we verified that a Lognormal function fit is a suitable function for fitting the empirical PDF for the drag parameter γ. Although the Lognormal is a long-tailed function, we can define the average value $\overline{\gamma} = 1. \times 10^{-7}\,\mathrm{km}^{-1}$. Also, the fit parameters appear to be quite close to those of Paper I. It is worth to note here that the Drag Based Ensemble Model (DBEM) implementation by Dumbović et al. (2018) uses as γ PDF a Gaussian function with $\mu = 1. \times 10^{-8}\,\mathrm{km}^{-1}$ and $\sigma = 0.5 \times 10^{-8}\,\mathrm{km}^{-1}$, that is 10 times smaller than the value we found. Similarly to the present study, in a more recent paper (Calogovic et al., 2021), they applied a reverse modeling procedure with the DBEM aimed to find optimal values for the DBM parameters. Their results showed that for the drag parameter a higher median value of about three times larger ($\gamma = 0.32 \times 10^{-7}\,\mathrm{km}^{-1}$) and an extended range of values ($\sigma = 0.7 \times 10^{-8}\,\mathrm{km}^{-1}$) are needed than the one used in the previous version of the DBEM, for lower MAE and ME in predictions. Interestingly, in the study of (Rollett et al., 2016), over the list of 21 ICME tracked by heliospheric imager, the average value of the fitted drag parameter is also generally larger than that of the previous models ($\approx 0.4 \times 10^{-7}\,\mathrm{km}^{-1}$, even excluding 4 cases which yielded unrealistic too-high values for γ), still following the
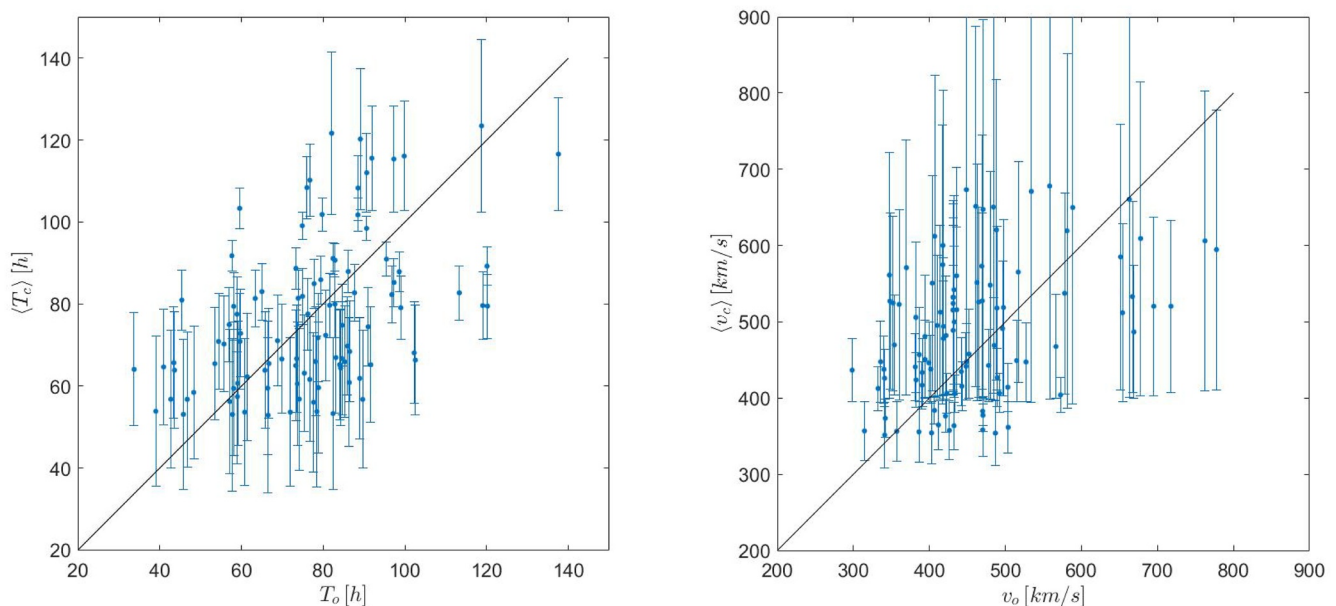
Figure 8. Plots of the computed mean Time of Arrival (ToA) versus the observed ToA (left) and of the computed mean Speed of Arrival (SoA) versus the observed SoA (right), obtained after applying the new PDBM distributions on the subset of events from (Paouris & Mavromichalaki, 2017a) list.

same trend. Furthermore, a recent paper by (Paouris, Čalogović, et al., 2021) found that the DBEM needs a larger value between 2.1 and 4.8 times larger of the drag parameter to model a set of CMEs. The large statistics of the inverted DBM parameters, allowed us also to try and separate the PDFs for those CMEs which are accelerated or decelerated by the solar wind. Although the statistics for the accelerated CMEs is much reduced, we found evidence that the γ PDFs are significantly different. In particular, the accelerated CMEs seem to experience, on average, a larger value of γ and to equalize more rapidly their speed to that of the solar wind.

We therefore introduce refinement of the PDBM with respect to that presented in Paper I, using different γ PDFs in case of accelerated or decelerated ICME. It is worth explaining why we chose to look for this separation. As the drag-based model describes the CME propagation taking as a reference the motion of a solid body in a fluid stream, a different value for the drag parameter may be expected if such body does not present the same shape to a fluid coming from the rear (accelerating CME) and to a fluid coming from the front (decelerated CME). We suggest that this may be the case, as ICMEs are typically depicted as curved flux tube, and our finding that the accelerated CMEs experience a higher drag than the decelerated ones is in accordance with such picture, where we expect a higher drag due to the fluid piling up in the rear of accelerating ICMEs, and a lower drag for decelerating ones, which undergo a smoother solar wind flow on their edges. Interestingly, some results from (Vrsnak, B. et al., 2008) may lead to similar conclusions. In this work, they investigated several relationships between the CME dynamics and the CME mass and found a correlation between the latter and the initial speed, concluding that since slower CMEs tend to have lower mass, larger values of γ are expected, as the drag parameter is related to the inverse of the CME mass (refer to equations in Vršnak et al. (2013)).

The updated method shows an improvement on predicted ToA average, both in the test against the initial database (Section 4.3) and in the validation against Paouris' CME list (Section 4.4), although the performances obtained with the old and new PDBM implementations are comparable and well within the error bars.

It is worth to note that the performance of our model is in agreement with the results from the investigation of (Vourlidas et al., 2019), which investigated the relation between the size of the database and the ToA MAE for several forecasting methods. Apparently, ToA MAE in the range 10–15 hr is the current limit on the performance obtained by almost all the methods for ICME forecasting, including numerical models. Our interpretation is that this limit is set by both the lack of knowledge about the actual state of the interplanetary medium and the large uncertainties on the CME initial properties.

This triggers two considerations. First, as long as the input data has such large errors, there will be probably little gain in tuning the model performance without taking into consideration those input errors. This, of course, affects also the new PDBM implementation we are proposing in this work. Second, it is important to test/validate the forecast procedures (especially those suited for real-time implementation) using a standard database of CMEs, in order to allow the comparison of the performance of models under the very same conditions. To this purpose, we think that data sets such as that presented in this work and that in (Paouris & Mavromichalaki, 2017a) will be of benefit to the CME modeling community when comparing the performances of different models and methods.

Lastly, it is worth to stress that the joint PDFs in Figure 3 show a non-linear correlation between γ and w: while we did not make use of this information in our work, the use of a joint PDF for the parameter extraction would reduce the parameter space by one degree of freedom. This approach definitely deserves further investigation, and it will be treated in future work.

Appendix A: Further Details on Methods

Within the Appendix we provide more information about the methods employed to build the CME database.

A1. Identifying the CME Source on the Solar Disk

To find the most probable CME source on the Sun, we employed a source finding algorithm that makes use of HEK (Heliophysics Event Knowledgebase-Hurlburt et al., 2010) to query which solar features (Active Regions, Solar Flares, Filaments Eruptions) that may have been the source of the CME, are within an area A and a time span Δt compatible with the CME launch parameters. The time span Δt is defined by an estimate of the time and duration of the CME liftoff obtained from LASCO images. The search area A is the whole solar sector defined by the CME POS angle and the angular width W of the CME if $W < 180°$ (normal and partial halo CMEs). In
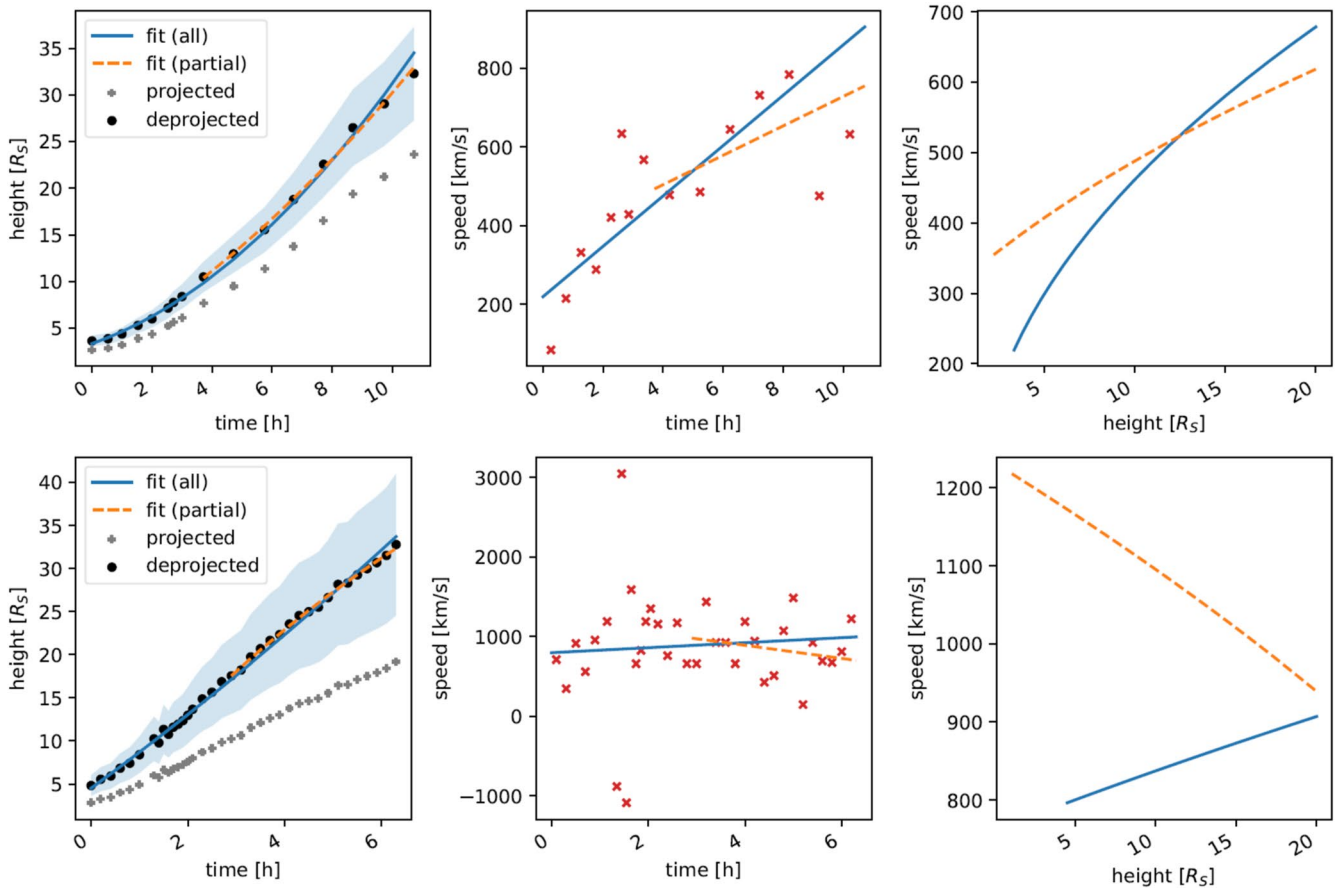
Figure A1. Illustration of the fitting procedure used on the Coronal Mass Ejections (CME) height versus time data. The two rows show examples for CME n.13 and n.179 from the data set. Left column: the original and deprojected data, with the shaded area indicating deprojection uncertainties. Two types of quadratic fits are shown: using all the data points (used for our database for robustness) and using only part of the data. Middle column: CME speed as determined from the fits and as sampled from the deprojected data. Right column: CME speed versus CME height, as determined from both types of fits. At 20 solar radii, they usually agree quite well.

the case $180° \leq W < 270°$ (half halo CME), the same sector is limited to 800″ from the disk center. In the case $W \geq 270°$ (full halo CME), A is the central part of the solar disk, within 600″ from the disk center. In the case multiple possible sources are retrieved by the query, the position of the CME source is the weighted average of the retrieved feature positions, with active regions weighting the larger between 0.1 and their associated R values (Schrijver, 2007), flares weighting 25 and filament ejections weighting 500. In those cases where no potential feature is retrieved by the query, the default source position is computed as the intersection between the CME POS angle vector and a circle centered on the disk center, with radius R^*, where $R^* = \frac{1}{3} R_{sun}$ in case of a full halo CME, $R^* = \frac{1}{2} R_{sun}$ in case of a half halo CME, $R^* = \frac{2}{3} R_{sun}$ in case of a partial halo CME, $R^* = R_{sun}$ in case the CME width is smaller than 90°. To the CME foot-point position we associate an error given by the larger between five″ and the standard deviation of the weighted averaging described above. We employed this method and these weighting values since they are those in use in the real-time CME detection and propagation services SWERTO (Berrilli et al., 2017) and IPS (Veettil et al., 2019) and have been set after an extensive test on a number of known CME-ICME counterparts.

A2. De-Projecting the CME Propagation Vector

After the identification of CME most likely source region, it is necessary to compute the radial speed v_r from the measured POS speed of the CME front. The procedure is based on Equation 1 in Gopalswamy et al. (2010), assuming a cone model for the CME shape (see Figure 1 in the same reference). The main difference is that we use the de-projection coefficient to obtain the de-projected position R_r from the POS position of the CME front,

instead of directly de-projecting the speed. By using such equation we can also compute the error associated to the radial position dR_r.

Once we have the data to plot a time-distance relation in the de-projected framework, we can obtain the radial speed by doing a linear or quadratic fit of the $r(t)$ relationship described by these data. The standard option is to fit a quadratic relationship, but when the number of measured position of the CME available for the fit is less than 9, a linear fit is used. Examples of quadratic fits are presented in Figure A1. Note that with a quadratic fit, the CME's acceleration is assumed to be constant and its speed is assumed to be a linear function. Because of this approximation, different velocities can be obtained depending on which part of the $r(t)$ data is used for the fit, as illustrated in the figure. In this paper, we have for robustness always used all available data points.

With the parameters from the fit, we are able to compute several quantities of interest for the CME liftoff. Namely, the time when the CME front reaches the $20R_{Sun}$ distance and its associated error; the CME v_r (@$20R_{Sun}$) and associated error; the possible CME front residual acceleration at $20R_{Sun}$ distance. Those values make part of the CME liftoff characteristics in the database: start date, de-projected speed, de-projected speed error, acceleration. Since the error on the arrival date is negligible, the error on the Start Date is reported as transit time error.

A3. Associating a Solar Wind Speed to Each CME

In order to propagate the ICME with an appropriate solar wind speed, for each event we have to hypothesize if the ICME interacted with a stream of slow (S) or fast (F) solar wind. It is well known that coronal holes are sources of fast solar wind streams (Krieger et al., 1973; Nolte et al., 1976), therefore we implemented an algorithm which discriminates the solar wind type by verifying if the CME source region is close to a coronal hole. A suitable algorithm queries the HEK (Heliophysics Event Knowledge) catalog for all the Coronal Holes present on the solar disk.

The time range queried starts from 4 hr before the estimated CME launch time to the CME launch time (considering the error).

As a consequence, we associate the event with fast (slow) solar wind if the CME source coordinates are close to (far from) any Coronal Hole retrieved by the query.

Data Availability Statement

The ICME catalog built for the analysis in Section 3, together with a tool for the data visualization and the module employed for running the PDBM simulations, can be downloaded from https://doi.org/10.5281/zenodo.5818470 (Napoletano et al., 2021).

References

Amerstorfer, T., Mööstl, C., Hess, P., Temmer, M., Mays, M., Reiss, M., et al. (2018). Ensemble prediction of a halo coronal mass ejection using heliospheric imagers. *Space Weather*, *16*(7), 784–801. https://doi.org/10.1029/2017sw001786

Berrilli, F., Casolino, M., Del Moro, D., Forte, R., Giovannelli, L., Martucci, M., et al. (2017). Swerto: A regional space weather service. *Proceedings of the International Astronomical Union*, *13*(S335), 348–351. https://doi.org/10.1017/s1743921318000054

Buzulukova, N. (2017). *Extreme events in geospace: Origins, predictability, and consequences*. Elsevier.

Calogovic, J., Dumbović, M., Sudar, D., Vršnak, B., Martinić, K., Temmer, M., & Veronig, A. M. (2021). Probabilistic drag-based ensemble model (DBEM) evaluation for heliospheric propagation of CMEs. *Solar Physics*, *296*(114). https://doi.org/10.1007/s11207-021-01859-5

Camporeale, E. (2019). The challenge of machine learning in space weather: Nowcasting and forecasting. *Space Weather*, *17*(8), 1166–1207. https://doi.org/10.1029/2018sw002061

Cargill, P. J. (2004). On the aerodynamic drag force acting on interplanetary coronal mass ejections. *Solar Physics*, *221*(1), 135–149. https://doi.org/10.1023/b:sola.0000033366.10725.a2

Cash, M., Biesecker, D., Pizzo, V., de Koning, C., Millward, G., Arge, C., et al. (2015). Ensemble modeling of the 23 July 2012 coronal mass ejection. *Space Weather*, *13*(10), 611–625. https://doi.org/10.1002/2015sw001232

Chen, J. (2017). Physics of erupting solar flux ropes: Coronal mass ejections (CMEs)—Recent advances in theory and observation. *Physics of Plasmas*, *24*(9), 090501. https://doi.org/10.1063/1.4993929

Chen, P. (2011). Coronal mass ejections: Models and their observational basis. *Living Reviews in Solar Physics*, *8*(1), 1. https://doi.org/10.12942/lrsp-2011-1

Daglis, I. A. (2001). Space storms, ring current and space-atmosphere coupling. In *Space storms and space weather hazards* (pp. 1–42). Springer. https://doi.org/10.1007/978-94-010-0983-6_1

Del Moro, D., Napoletano, G., Forte, R., Giovannelli, L., Pietropaolo, E., & Berrilli, F. (2019). Forecasting the 2018 February 12th CME propagation with the P-DBM model: A fast warning procedure. *Annals of Geophysics*, *62*(4), 456. https://doi.org/10.4401/ag-7750

Dumbovic, M., Calogovic, J., Martinic, K., Vrsnak, B., Sudar, D., Temmer, M., & Veronig, A. (2021). *Drag-based model (DBM) tools for forecast of coronal mass ejection arrival time and speed*. arXiv preprint arXiv:2103.14292.

Dumbović, M., Čalogović, J., Vršnak, B., Temmer, M., Mays, M. L., Veronig, A., & Piantschitsch, I. (2018). The drag-based ensemble model (DBEM) for coronal mass ejection propagation. *The Astrophysical Journal*, *854*(2), 180.

Emmons, D., Acebal, A., Pulkkinen, A., Taktakishvili, A., MacNeice, P., & Odstrcil, D. (2013). Ensemble forecasting of coronal mass ejections using the WSA-ENLIL with CONED model. *Space Weather*, *11*(3), 95–106. https://doi.org/10.1002/swe.20019

Gopalswamy, N., Lara, A., Yashiro, S., Kaiser, M. L., & Howard, R. A. (2001). Predicting the 1-au arrival times of coronal mass ejections. *Journal of Geophysical Research*, *106*(A12), 29207–29217. https://doi.org/10.1029/2001ja000177

Gopalswamy, N., Yashiro, S., Michalek, G., Xie, H., Makela, P., Vourlidas, A., & Howard, R. A. (2010). A catalog of halo coronal mass ejections from SOHO. *Sun and Geosphere*, *5*, 7–16.

Hurlburt, N., Cheung, M., Schrijver, C., Chang, L., Freeland, S., Green, S., et al. (2010). Heliophysics event knowledge base for the solar dynamics observatory (SDO) and beyond. In *The solar dynamics observatory* (pp. 67–78). Springer. https://doi.org/10.1007/978-1-4614-3673-7_5

Iwai, K., Shiota, D., Tokumaru, M., Fujiki, K., Den, M., & Kubo, Y. (2019). Development of a coronal mass ejection arrival time forecasting system using interplanetary scintillation observations. *Earth, Planets and Space*, *71*(1), 39. https://doi.org/10.1186/s40623-019-1019-5

Kay, C., & Gopalswamy, N. (2018). The effects of uncertainty in initial CME input parameters on deflection, rotation, b z, and arrival time predictions. *Journal of Geophysical Research: Space Physics*, *123*(9), 7220–7240. https://doi.org/10.1029/2018ja025780

Kay, C., Mays, M., & Verbeke, C. (2020). Identifying critical input parameters for improving drag-based cme arrival time predictions. *Space Weather*, *18*(1), e2019SW002382. https://doi.org/10.1029/2019sw002382

Kilpua, E., Koskinen, H. E., & Pulkkinen, T. I. (2017). Coronal mass ejections and their sheath regions in interplanetary space. *Living Reviews in Solar Physics*, *14*(1), 5. https://doi.org/10.1007/s41116-017-0009-6

Kilpua, E., Mierla, M., Rodriguez, L., Zhukov, A., Srivastava, N., & West, M. (2012). Estimating travel times of coronal mass ejections to 1 au using multi-spacecraft coronagraph data. *Solar Physics*, *279*(2), 477–496. https://doi.org/10.1007/s11207-012-0005-x

Krieger, A. S., Timothy, A. F., & Roelof, E. C. (1973). A coronal hole and its identification as the source of a high velocity solar wind stream. *Solar Physics*, *29*(2), 505–525. https://doi.org/10.1007/BF00150828

Lamy, P. L., Floyd, O., Boclet, B., Wojak, J., Gilardy, H., & Barlyaeva, T. (2019). Coronal mass ejections over solar cycles 23 and 24. *Space Science Reviews*, *215*(5), 39. https://doi.org/10.1007/s11214-019-0605-y

Martens, P., Attrill, G., Davey, A., Engell, A., Farid, S., Grigis, P., et al. (2012). Computer vision for the solar dynamics observatory (SDO). *Solar Physics*, *275*(1–2), 79–113. https://doi.org/10.1007/s11207-010-9697-y

Mays, M., Taktakishvili, A., Pulkkinen, A., MacNeice, P., Rastätter, L., Odstrcil, D., et al. (2015). Ensemble modeling of CMEs using the WSA–ENLIL+ cone model. *Solar Physics*, *290*(6), 1775–1814. https://doi.org/10.1007/s11207-015-0692-1

Mumford, S., Freij, N., Christe, S., Ireland, J., Mayer, F., Hughitt, V., et al. (2020). SunPy: A python package for solar physics. *Journal of Open Source Software*, *5*(46), 1832. https://doi.org/10.21105/joss.01832

Napoletano, G., Foldes, R., Del Moro, D., De Gasperis, G., Tiwari, A. K., Teunissen, J., & Camporeale, E. (2021). Earth-directed coronal mass ejections from 1996 to 2018. Version 2.0 [Data set] Zenodo. https://doi.org/10.5281/zenodo.5818470

Napoletano, G., Forte, R., Del Moro, D., Pietropaolo, E., Giovannelli, L., & Berrilli, F. (2018). A probabilistic approach to the drag-based model. *Journal of Space Weather and Space Climate*, *8*, A11. https://doi.org/10.1051/swsc/2018003

Nolte, J. T., Krieger, A. S., Timothy, A. F., Gold, R. E., Roelof, E. C., Vaiana, G., et al. (1976). Coronal holes as sources of solar wind. *Solar Physics*, *46*(2), 303–322. https://doi.org/10.1007/BF00149859

Odstrcil, D. (2003). Modeling 3-D solar wind structure. *Advances in Space Research*, *32*(4), 497–506. https://doi.org/10.1016/s0273-1177(03)00332-6

Owens, M., Lang, M., Barnard, L., Riley, P., Ben-Nun, M., Scott, C. J., & Gonzi, S. (2020). A computationally efficient, time-dependent model of the solar wind for use as a surrogate to three-dimensional numerical magnetohydrodynamic simulations. *Solar Physics*, *295*(3), 1–17. https://doi.org/10.1007/s11207-020-01605-3

Paouris, E., Čalogović, J., Dumbović, M., Mays, M. L., Vourlidas, A., Papaioannou, A., et al. (2021). Propagating conditions and the time of ICME arrival: A comparison of the effective acceleration model with ENLIL and DBEM models. *Solar Physics*, *296*(1), 12. https://doi.org/10.1007/s11207-020-01747-4

Paouris, E., & Mavromichalaki, H. (2017a). Interplanetary coronal mass ejections resulting from earth-directed CMEs using SOHO and ACE combined data during solar cycle 23. *Solar Physics*, *292*(2), 30. https://doi.org/10.1007/s11207-017-1050-2

Paouris, E., & Mavromichalaki, H. (2017b). Effective acceleration model for the arrival time of interplanetary shocks driven by coronal mass ejections. *Solar Physics*, *292*(12). https://doi.org/10.1007/s11207-017-1212-2

Paouris, E., Vourlidas, A., Papaioannou, A., & Anastasiadis, A. (2021). Assessing the projection correction of coronal mass ejection speeds on time-of-arrival prediction performance using the effective acceleration model. *Space Weather*, *19*(2), e2020SW002617. https://doi.org/10.1029/2020sw002617

Piersanti, M., De Michelis, P., Del Moro, D., Tozzi, R., Pezzopane, M., Consolini, G., et al. (2020). From the sun to earth: Effects of the 25 August 2018 geomagnetic storm. *Annales Geophysicae*, *38*(3), 703–724. https://doi.org/10.5194/angeo-38-703-2020

Pomoell, J., & Poedts, S. (2018). EUHFORIA: European heliospheric forecasting information asset. *Journal of Space Weather and Space Climate*, *8*, A35. https://doi.org/10.1051/swsc/2018020

Pulkkinen, T. (2007). Space weather: Terrestrial perspective. *Living Reviews in Solar Physics*, *4*(1), 1. https://doi.org/10.12942/lrsp-2007-1

Richardson, I. G., & Cane, H. V. (2010). Near-earth interplanetary coronal mass ejections during solar cycle 23 (1996–2009): Catalog and summary of properties. *Solar Physics*, *264*(1), 189–237. https://doi.org/10.1007/s11207-010-9568-6

Rollett, T., Möstl, C., Isavnin, A., Davies, J. A., Kubicka, M., Amerstorfer, U. V., & Harrison, R. A. (2016). ElEvoHI: A novel CME prediction tool for heliospheric imaging combining an elliptical front with drag-based model fitting. *The Astrophysical Journal*, *824*(2), 131. https://doi.org/10.3847/0004-637x/824/2/131

Rollett, T., Möstl, C., Temmer, M., Frahm, R. A., Davies, J. A., Veronig, A. M., et al. (2014). Combined multipoint remote and in situ observations of the asymmetric evolution of a fast solar coronal mass ejection. *The Astrophysical Journal*, *790*(1), L6. https://doi.org/10.1088/2041-8205/790/1/l6

Schrijver, C. J. (2007). A characteristic magnetic field pattern associated with all major solar flares and its use in flare forecasting. *The Astrophysical Journal Letters*, *655*(2), L117–L120. https://doi.org/10.1086/511857

Schrijver, C. J., & Siscoe, G. L. (2010). *Heliophysics: Space storms and radiation: Causes and effects*. Cambridge University Press.

Schwenn, R. (2006). Space weather: The solar perspective. *Living Reviews in Solar Physics*, *3*(1), 2. https://doi.org/10.12942/lrsp-2006-2

Schwenn, R., Dal Lago, A., Huttunen, E., & Gonzalez, W. D. (2005). The association of coronal mass ejections with their effects near the Earth. *Annales Geophysicae*, *23*(3), 1033–1059. https://doi.org/10.5194/angeo-23-1033-2005

Temmer, M., Vršnak, B., Rollett, T., Bein, B., de Koning, C. A., Liu, Y., et al. (2012). Characteristics of kinematics of a coronal mass ejection during the 2010 August 1 CME–CME interaction event. *The Astrophysical Journal*, *749*(1), 57. https://doi.org/10.1088/0004-637x/749/1/57

The SunPy Community, Barnes, W. T., Bobra, M. G., Christe, S. D., Freij, N., Hayes, L. A., Dang, T. K., et al. (2020). The SunPy project: Open source development and status of the version 1.0 core package. *The Astrophysical Journal*, *890*, 68. https://doi.org/10.3847/1538-4357/ab4f7a

Tsurutani, B. T., Gonzalez, W. D., Tang, F., Akasofu, S. I., & Smith, E. J. (1988). Origin of interplanetary southward magnetic fields responsible for major magnetic storms near solar maximum (1978–1979). *Journal of Geophysical Research*, *93*(A8), 8519–8531. https://doi.org/10.1029/ja093ia08p08519

Veettil, S. V., Cesaroni, C., Aquino, M., De Franceschi, G., Berrili, F., Rodriguez, F., et al. (2019). The ionosphere prediction service prototype for GNSS users. *Journal of Space Weather and Space Climate*, *9*, A41. https://doi.org/10.1051/swsc/2019038

Verbeke, C., Mays, M., Temmer, M., Bingham, S., Steenburgh, R., Dumbović, M., et al. (2019). Benchmarking CME arrival time and impact: Progress on metadata, metrics, and events. *Space Weather*, *17*(1), 6–26. https://doi.org/10.1029/2018sw002046

Vourlidas, A., Patsourakos, S., & Savani, N. (2019). Predicting the geoeffective properties of coronal mass ejections: Current status, open issues and path forward. *Philosophical Transactions of the Royal Society A*, *377*(2148), 20180096. https://doi.org/10.1098/rsta.2018.0096

Vršnak, B., Temmer, M., Žic, T., Taktakishvili, A., Dumbović, M., Möstl, C., & Odstrčil, D. (2014). Heliospheric propagation of coronal mass ejections: Comparison of numerical WSA-ENLIL+ cone model and analytical drag-based model. *The Astrophysical Journal-Supplement Series*, *213*(2), 21.

Vrsnak, B., Vrbanec, D., & Calogoviˊc, J. (2008). Dynamics of coronal mass ejections-The mass-scaling of the aerodynamic drag. *A&A*, *490*(2), 811–815. https://doi.org/10.1051/0004-6361:200810215

Vršnak, B., Žic, T., Vrbanec, D., Temmer, M., Rollett, T., Möstl, C., et al. (2013). Propagation of interplanetary coronal mass ejections: The drag-based model. *Solar Physics*, *285*(1–2), 295–315.

Webb, D. F., & Howard, T. A. (2012). Coronal mass ejections: Observations. *Living Reviews in Solar Physics*, *9*(1), 3. https://doi.org/10.12942/lrsp-2012-3

Žic, T., Vršnak, B., & Temmer, M. (2015). Heliospheric propagation OF coronal mass ejections: Drag-based model fitting. *The Astrophysical Journal - Supplement Series*, *218*(2), 32. https://doi.org/10.1088/0067-0049/218/2/32