The Drag-based Ensemble Model (DBEM) for Coronal Mass Ejection Propagation

Mateja Dumbović, Jaša Čalogović, Bojan Vršnak, Manuela Temmer, M. Leila Mays, Astrid Veronig, and Isabell Piantschitsch

1 Institute of Physics, University of Graz, Universitätsplatz 5, A-8010 Graz, Austria; mateja.dumbovic@uni-graz.at
2 Hvar Observatory, Faculty of Geodesy, University of Zagreb, Kačićeva 26, HR-10000, Zagreb, Croatia
3 NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
4 Kanzelhöhe Observatory for Solar and Environmental Research, University of Graz, Universitätsplatz 5, A-8010 Graz, Austria

Received 2017 October 25; revised 2018 January 16; accepted 2018 January 22; published 2018 February 26

Abstract

The drag-based model for heliospheric propagation of coronal mass ejections (CMEs) is a widely used analytical model that can predict CME arrival time and speed at a given heliospheric location. It is based on the assumption that the propagation of CMEs in interplanetary space is solely under the influence of magnetohydrodynamical drag, where CME propagation is determined based on CME initial properties as well as the properties of the ambient solar wind. We present an upgraded version, the drag-based ensemble model (DBEM), that covers ensemble modeling to produce a distribution of possible ICME arrival times and speeds. Multiple runs using uncertainty ranges for the input values can be performed in almost real-time, within a few minutes. This allows us to define the most likely ICME arrival times and speeds, quantify prediction uncertainties, and determine forecast confidence. The performance of the DBEM is evaluated and compared to that of ensemble WSA-ENLIL+Cone model (ENLIL) using the same sample of events. It is found that the mean error is ME = −9.7 hr, mean absolute error MAE = 14.3 hr, and root mean square error RMSE = 16.7 hr, which is somewhat higher than, but comparable to ENLIL errors (ME = −6.1 hr, MAE = 12.8 hr and RMSE = 14.4 hr). Overall, DBEM and ENLIL show a similar performance. Furthermore, we find that in both models fast CMEs are predicted to arrive earlier than observed, most likely owing to the physical limitations of models, but possibly also related to an overestimation of the CME initial speed for fast CMEs.

Key words: magnetohydrodynamics (MHD) – methods: analytical – methods: statistical – solar–terrestrial relations – solar wind – Sun: coronal mass ejections (CMEs)

1. Introduction

Because coronal mass ejections (CMEs) dominate space weather effects and entail potentially harmful impacts on Earth, the forecasting of CMEs is one of the major challenges for space weather forecasts. Therefore, in recent years many CMEs and their associated shock propagation models have been developed by research groups around the globe, and space weather forecast centers regularly implement some of these propagation models to warn of the possible arrival of potentially threatening CMEs (e.g., the Space Weather Prediction Center of the National Oceanic and Atmospheric Administration, SWPC/NOAA, in the USA, the Met Office in the UK, and the Solar Influences Data Analysis Center, SIDC, in Belgium).

The propagation models differ based on the input, approach, assumptions, and complexity (for an overview see, e.g., Zhao & Dryer 2014, and references therein) and vary from simple empirical models (e.g., Gopalswamy et al. 2001), neural network models (e.g., Sudar et al. 2016), analytical drag-based models with different geometries (e.g., Vršnak et al. 2014; Rollett et al. 2016), various kinematic shock propagation models (e.g., Dryer et al. 2001; Zhao et al. 2016; Takahashi & Shibata 2017), to complex numerical 3D MHD models such as the H3DMHD model (Wu et al. 2011) or WSA-ENLIL+Cone model (Odstrcil et al. 2004). Despite these differences, in general most of the models show a surprisingly comparable performance, where the prediction errors are mostly found within the 24 hr interval and the mean absolute error is ~10 hr (e.g., Gopalswamy et al. 2001; Li et al. 2008; Vršnak et al. 2014; Mays et al. 2015; Sudar et al. 2016; Wold et al. 2017).

The drag-based models are based on the concept of MHD drag, which, unlike the kinetic drag effect in a fluid, is presumed to be caused primarily by the emission of MHD waves in the collisionless solar wind environment and acts to adjust the CME speed to the ambient solar wind (Cargill et al. 1996). The concept of drag relies on the observational fact that slow CMEs accelerate and fast CMEs decelerate (e.g., Sheeley et al. 1999; Gopalswamy et al. 2000) and is supported by numerous studies (e.g., Vršnak & Žic 2007; Temmer et al. 2011; Liu et al. 2013; Hess & Zhang 2014; Sachdeva et al. 2015, and references therein). Vršnak & Žic (2007) proposed that the equation describing the aerodynamic drag can be utilized to establish a simple kinematical drag-based model (DBM) for CME propagation, which has since been developed by implementing the cone geometry of a CME and performing a parametric analysis to empirically determine the drag parameter $\gamma$ (Vršnak et al. 2013, 2014; Žic et al. 2015). DBM is available at the Hvar Observatory website as an online tool,4 is one of the European Space Agency (ESA) space situational awareness (SSA) products,6 is one of the models available at the Community Coordinated Modeling Center (CCMC),7 and is incorporated into the automatic COMESE system (Crosby et al. 2012; Dumbović et al. 2017). The performance of DBM was shown to be comparable to that of other propagation models (Vršnak et al. 2014) and was found to agree well with the WSA-ENLIL+Cone model.

4 http://oh.geof.unizg.hr/DBM/dbm.php
6 http://swe.ssa.esa.int/heliospheric-weather
7 https://ccmc.gsfc.nasa.gov
model, which is one of the most extensively used CME propagation models in space weather operations worldwide.

The similarities in the performances of very different propagation models indicates that the major drawback for more accurately forecasting CME arrival times and impact speeds is the lack of reliable observation-based input. In order to take into account the errors and uncertainties in the CME measurements that are used as model input and to quantify the resulting uncertainties in the model predictions, ensemble forecasting is widely used. Recently, Mays et al. (2015) used an ensemble modeling approach to evaluate the sensitivity of the WSA-ENLIL+Cone model (hereafter ENLIL) simulations to initial CME parameters and provide a probabilistic forecasting of CME arrival time. Ensemble modeling takes into account the variability of observation-based model input by making an ensemble, i.e., sets of n CME observations to calculate a distribution of predictions and forecast the confidence in the likelihood of the CME arrival. We use the ensemble approach in DBM and present the newly developed drag-based ensemble model (DBEM). The model is evaluated with the same data set that Mays et al. (2015) used, to make it comparable to ENLIL ensemble results.

2. Data and Method

DBEM is based on DBM, with an assumed cone geometry for the CME, where the leading edge is initially a semicircle, spanning over the full angular width of the CME, and flattens as it evolves in time (described in Žic et al. 2015). The model assumes a constant solar wind speed, w, and drag parameter, γ, which is in general valid for distances beyond $R > 15 R_{\text{SO}}$, where the CME moves in an isotropic solar wind spreading out at a constant speed and the fall-off of the ambient density is at the same rate as the CME expansion (see Vršnak et al. 2013; Žic et al. 2015). As this assumption is clearly not valid for CME–CME interaction events (Temmer et al. 2012), we do not consider such interaction events in our study. We note that the constant w and γ assumption can be generally valid even for CMEs moving in high-speed streams, assuming they encounter high-speed streams relatively close to the Sun.

The input parameters derived from observations and used as input for the DBEM are the CME speed, half-width, and propagation direction (longitude) defined at a certain distance/time from the Sun. Since this is a 2D model that operates in the ecliptic plane, DBEM does not use CME latitude as an input. The solar wind speed (the radial component in the ecliptic plane) and the drag parameter, γ, complete the input values. As a first step, for a single CME, we use an ensemble of n measurements of the same CME as Mays et al. (2015; see Section 2.1). Next, the variability of DBM parameters (solar wind speed, w and drag parameter, γ) is taken into account, where m synthetic values of both w and γ are produced (see Section 2.2). These synthetic values are combined with an ensemble of n CME measurements, giving a final ensemble of $n \cdot m^2$ members as an input, which, after $n \cdot m^2$ runs, produces a distribution of $n \cdot m^2$ calculated CME transit times and arrival speeds.

2.1. CME Initial Parameters

We use the sample compiled and analyzed by Mays et al. (2015), which consists of 35 CMEs and associated interplanetary CMEs (ICMEs, if detected at Earth) in the time period 2013 January to 2014 July. All CME measurements and simulation summary results are available at https://iswa.ccmc.gsfc.nasa.gov/ENSEMBLE/. The CME initial parameters are determined using the Stereoscopic CME Analysis Tool (StereoCAT) developed by CCMC. StereoCAT tracks specific CME features, based on triangulation of transient CME features manually identified using two different coronagraph fields of view. From this, the 3D speed and position are derived for a CME, as well as its (projected) width. To gather an ensemble of measurements, the CME leading edge height was measured at two different times in each coronagraph image for two different coronagraph viewpoints and then the procedure was repeated $k$ times to obtain an ensemble of $n = k^2$ CME measurements (for details see Mays et al. 2015). Each ensemble member has a specific set of initial CME parameters—speed, width, longitude, and latitude. Given that a typical run of the whole ensemble using ENLIL simulations is 80–130 minutes (depending on the computing power), for each event an optimal spread of input parameters was selected ($n = 12, 16, 24, 36, 48$). Note that the input is more suitable for the 3D ENLIL WSA+cone model, which uses both longitude and latitude as positional parameters and 3D velocity as input speed, whereas for DBEM the radial component of velocity in the ecliptic plane would be a more suitable input.

We refine the sample compiled by Mays et al. (2015), excluding events with CME–CME interaction and events where in situ arrival times could not be determined exactly and unambiguously. The resulting sample consists of 25 CMEs, and for each event we use as ensemble model input the derived speed, half-width, longitude, and start time. The start time corresponds to the starting distance, which is restrained by the ENLIL inner boundary corresponding to $R = 21.5 R_{\text{SO}}$ and is also suitable for DBM due to the assumption of constant $w$ and $\gamma$, and furthermore because past that distance, the drag is the dominant force governing the propagation of ICMEs for a large subset of CMEs (Sachdeva et al. 2015, 2017).

2.2. Solar Wind Speed and Drag Parameter $\gamma$

As input values for the solar wind speed, $w$, and drag parameter, $\gamma$, we do not use measured values, but empirical values derived in previous studies by Vršnak et al. (2013, 2014). Therefore, to include their uncertainty in DBEM in the same way that we have included the uncertainty of the CME input (CME ensembles), synthetic values are needed. Synthetic values of $w$ and $\gamma$ for a specific event can be produced assuming that the real measurements of these parameters follow a normal distribution fully defined by the following expression:

$$x = \mu \pm \Delta x,$$

where $\mu$ is the mean of the normal distribution and $\Delta x = 3 \cdot \sigma$ defines a range where 99.7% measurements are found ($\sigma$ is the standard deviation). The normal distribution for a random variable $x$, $f(x)$, is defined for a specific $\mu$ and $\sigma$, where $w$ and $\gamma$ are treated as random variables described by a normal distribution. In Figures 1(a) and (b) examples of normal distributions are shown for $w = (350 \pm 50) \text{ km s}^{-1}$ and $\gamma = (0.1 \pm 0.05) \cdot 10^{-7} \text{ km}^{-1}$, respectively. Based on
our assumption, these are the distributions we would obtain by making real measurements of \( w \) and \( \gamma \), where their real values are found in the intervals \( w = (350 \pm 50) \text{ km s}^{-1} \) and \( \gamma = (0.1 \pm 0.05) \cdot 10^{-7} \text{ km}^{-1} \), respectively.

Using the substitution \( z = (x - \mu)/\sigma \), a corresponding standard normal distribution (SND), \( f(z) \), is obtained, which is normalized with respect to the original distribution so that \( \mu = 0 \) and \( \sigma = 1 \) (shown in Figure 1(c)). The cumulative SND, which is the probability that \( z \) will take a value less than or equal to some \( z_0 \) (i.e., gives the area under \( f(z) \) from \(-\infty\) to \( z_0 \)), can be written as

\[
\Phi(z) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{z}{\sqrt{2}} \right) \right],
\]

where the \( \text{erf} (z) \) is the Gauss error function defined as

\[
\text{erf} (z) = \frac{2}{\sqrt{\pi}} \int_{0}^{z} e^{-z^2} dz.
\]

\( \Phi(z) \) is a continuous function defined on an interval \([0, 1]\). However, if multiplied with a normalization factor \( m - 1 \), where \( m \) is the number of synthetic values we wish to produce, we obtain a new continuous function \( \Phi^*(z) \) defined on an interval \([0, m-1]\), which can be used to obtain \( m \) different values of \( z \):

\[
z_i = -\sqrt{2} \cdot \text{erf}^{-1} \left( 1 - 2 \cdot \frac{\Phi^*_i(z)}{m - 1} \right),
\]

\[
\Phi^*_i(z) = 0, 1, 2, \ldots, m - 2, m - 1.
\]  

This is graphically represented in Figure 1(d) in an example where \( m = 6 \). Each \( z_i \) corresponds to a certain \( x_i \) based on the substitution \( z_i = (x_i - \mu)/\sigma \). Therefore, for a given number \( m \) and a parameter defined according to Equation (1), \( m \) synthetic values of the parameter can be produced based on Equation (4).

An example is given in Figures 2(a) and (b), where \( m = 25 \) synthetic values are shown against the normal distributions for \( w = (350 \pm 50) \text{ km s}^{-1} \) and \( \gamma = (0.1 \pm 0.05) \cdot 10^{-7} \text{ km}^{-1} \), respectively.

Based on Figure 2 it is clear that the synthetic values will reflect the normal distribution closer if larger \( m \) is chosen. However, very large \( m \) will increase the number of runs and consequently increase computing time. Therefore, the optimum \( m \) needs to be determined. For that purpose, we randomly selected a CME from our list that has a total of 48 ensemble members (2014 February 12). We then randomly selected 3, 5, 8, 12, 24, and 36 ensemble members to simulate the respective ensemble sizes of the same CME (different ensemble sizes are
denoted with $n$). For each of the ensembles we made model runs using different numbers of synthetic values, $m = 3, 5, 7, 10, 15, 20, 30, 50, 100$, and derived distributions of transit times, $T_T(\text{hr})$, and arrival speeds, $v(\text{km s}^{-1})$, from which we calculated the median and 95% confidence interval. The model runs include both the variability of the CME input ($n$) and the variability of the model parameters ($m$). In order to observe how the values of distribution median and 95% confidence interval change with $n$ and $m$, we normalized for each $n$ the difference of the $m$th median from the median

Figure 2. Normal distributions for (a) $w = (350 \pm 50)$, km s$^{-1}$ and (b) $\gamma = (0.1 \pm 0.05) \cdot 10^{-7}$ km$^{-1}$ and corresponding synthetic values (red circles).

Figure 3. Variability of the distribution median for (a) CME transit time and (b) arrival speed, and of the 95% confidence interval for (c) CME transit time and (d) arrival speed, depending on the number of CME ensemble members, $n$, and the number of solar wind speed and $\gamma$ parameter synthetic values, $m$. The variabilities are normalized to values corresponding to $m = 100$ (for each CME ensemble separately). The red dashed line marks the optimal value $m = 15$ (for details see the main text).
corresponding to a value of 100 in the following way:

\[
dX(n, m) = \frac{|X(n, m) - X(n, 100)|}{X(n, 100)},
\]

where \(X(n, m)\) is the median of \(TT\) and \(v\) for CME transit time and arrival speed, respectively. We applied the same equation to the 95% confidence interval as well, where for each \(n\) the difference of the \(m\)th 95% confidence interval is normalized to the 95% confidence interval corresponding to \(m = 100\) according to Equation (5), with \(X(n, m)\) being \(TT\) and \(v\) range for CME transit time and arrival speed, respectively. This is shown in Figure 3, where it can be seen that the variability of the median for both arrival speed and transit time is quite small (<1% for arrival speed, <3% for transit time) and decreases with both \(n\) and \(m\). The variability of the 95% confidence interval is much larger (it can go up to 30% for transit time and even 60% for arrival speed), but also decreases quickly with increasing \(n\) and \(m\). Both the median and the 95% confidence interval converge quite fast toward the value corresponding to \(m = 100\). At \(m = 15\) the variability of the median is already below 0.5% for all CME ensemble sizes \(n\) for both \(TT\) and \(v\), whereas the variability of the 95% confidence interval is below 5% for \(TT\) and around 10% for \(v\). We note that the typical prediction errors of the CME transit time are \(\sim 10\) hr (as described in Section 1), thus a 5% variation in a confidence interval would correspond to less than \(\sim 1\) hr and a 0.5% variation of the median would be of the order of \(\sim 1\) minute. Therefore, as an optimal value we choose \(m = 15\). For an ensemble of \(n = 48\) CME inputs, with this optimal value of \(m = 15\) synthetic \(w\) and \(\gamma\) values the total number of model runs is \(n \cdot m^2 = 10800\), resulting in a computational time of several minutes on an average PC.

For simplicity we use the same value of mean solar wind speed and \(\gamma\) parameter for all events in the sample. Due to the weak solar activity throughout the last solar cycle, which was reflected on the solar wind and interplanetary magnetic field (e.g., McComas et al. 2013), we use \(w = (350 \pm 50)\) km s\(^{-1}\). DBM tracks the leading edge of the ICME ejecta, whereas ENLIL tracks the shock front. However, it was shown that there is in general good agreement between the two, with a convenient selection of the \(\gamma\) parameter (\(\gamma = 0.1 \cdot 10^{-7}\) km\(^{-1}\)) (see Vršnak et al. 2014). Therefore, in order to estimate shock arrival with DBEM we select \(\gamma = (0.1 \pm 0.05) \cdot 10^{-7}\) km\(^{-1}\). With these values of \(w\) and \(\gamma\) we determine \(m = 15\) synthetic values of \(w\) and \(\gamma\) with the procedure described above. Note that the simplification of using average solar wind conditions results in a less realistic solar wind background than the one used by ENLIL.

### 3. Results and Discussion

For each ensemble member, i.e., individual run, DBEM calculates whether or not the CME will hit or miss the Earth. The corresponding condition can be written as \(\omega \geq |\phi|\), where \(\omega\) is the CME half-width and \(\phi\) is the CME source position longitude. For the whole ensemble, DBEM calculates the probability of the arrival as \(p = n_{\text{hit}}/n_{\text{tot}}\), where \(n_{\text{hit}}\) is the number of ensemble members that are calculated to hit Earth and \(n_{\text{tot}}\) is the total number of all ensemble members. The probability of arrival for a CME is displayed as a pie chart, as shown in the upper right panel of Figure 4, where DBEM results for the example CME 2013 August 30 are given. As can be seen in Figure 4, the probability of arrival for this event, as calculated by DBEM is 0.92 (91.7%, red part of the pie chart). The CME input for this event consists of 48 ensemble members with a start time ranging 05:59 UT to 06:30 UT, CME speed range of 810–1012 km s\(^{-1}\), a longitude range of E55°–E35°, and a half-width range of 41°–63°. This is supplemented with 15 synthetic values of solar wind speed in the range 300–400 km s\(^{-1}\) and 15 synthetic values of the drag parameter \(\gamma\) in the range 0.05–0.15 \(\cdot 10^{-7}\) km\(^{-1}\). Therefore, the size of the whole ensemble for this example is 10800, i.e., the results are based on 10800 DBM runs. A table of the first eight ensemble members (input lines) is given in the top left panel as a quicklook visualization of the input.

The bottom panels in Figure 4 show transit time, \(TT\), and arrival speed, \(v\), distributions, calculated only based on the runs for ensemble members calculated to hit Earth (the red part of the pie chart). The expected range is given by the 95% confidence interval. The most likely arrival date and time and corresponding range are given on the top of Figure 4 and are also calculated only based on ensemble members predicted to hit Earth. The output also provides a quick performance record—the number of runs and the computing time. The example event shown in Figure 4 was run on an average PC. Depending on the number of CPUs and their speeds, the current DBEM version can make a thousand (on single thread/CPUs) or more runs per second and can run the whole ensemble in several minutes, which is very fast compared to the ENLIL runtime (>1 hr on a high-performance machine, see Section 2.1). We perform the DBEM forecast evaluation for a set of 25 events, and when possible, compare the outcome to that of ENLIL. Forecast evaluation is based on DBEM and ENLIL calculations of \(TT\), DBM calculation of \(v\), and in situ observations of \(TT\) and \(v\) presented in Table 1 (for additional information on the median values of the CME input we refer the reader to Table 1 of Mays et al. 2015). We use the peak value of the in situ speed to be consistent in all events, due to the fact that sometimes only shock/sheath is encountered, sometimes only ICME magnetic structure and sometimes both.

The probability of arrival, \(p_i\), is binned into a categorical yes/no forecast using a \(p_i < 15\%\) criterion for correct rejection, the same as what was used in Mays et al. (2015). Thus, a \(2 \times 2\) contingency table for a binary event can be constructed and used to perform forecast evaluation (see e.g., Jolliffe & Stephenson 2003), where CME arrival is regarded as an event, and the event forecast as well as the event observation can have two outcomes, yes or no. There are four possible combinations of forecast and observation outcomes: a “hit” where the event was forecasted to hit Earth and observed, a “miss” where the event was not forecasted to hit Earth but was observed, a “false alarm” where the event was forecasted to hit Earth but was not observed and a “correct rejection,” where the event was neither forecasted nor observed. Following the procedure by Mays et al. (2015), if \(p_i < 15\%\) and there were no in situ signatures, the event (CME arrival) is considered a “correct rejection”; if \(p_i > 15\%\) and there were no in situ signatures, the event is considered as a “false alarm”; if \(p_i > 15\%\) and there are in situ signatures, the event is considered as a “hit”; and finally, if \(p_i < 15\%\) and there are in situ signatures, the event is considered a “miss.”
The number of hits, misses, false alarms, and correct rejections for our sample is given in Table 2(a) for DBEM and ENLIL, and the outcome using different evaluation measures is given in Table 2(b). Compared to ENLIL, DBEM has out of 25 events 1 false alarm more and 1 correct rejection less than ENLIL, which results in a somewhat lower performance. In the final row of Table 2(b) a Brier score (BS) is given, which quantifies the probability forecast errors:

\[
BS = \frac{1}{N} \sum_{i=1}^{N} (p_i - o_i)^2,
\]

where \(N\) is the total number of events, \(p_i\) is the forecast probability that event \(i\) will occur, and \(o_i\) takes a value 0 if event \(i\) did not occur and 1 if the event \(i\) did occur. For a perfect forecast \(BS = 0\) and we can see that both DBEM and ENLIL are not too far away from this value.

In Table 2(c) transit time errors are given for DBEM and ENLIL, calculated for “hit” events. It can be seen that DBEM has somewhat larger errors than ENLIL. One of the reasons is that the CME input used is more suitable for ENLIL than DBEM. The optimal CME input for ENLIL is full 3D information on the CME as given in the sample (3D velocity, longitude, and latitude), while for DBEM optimal input would be the radial velocity in the ecliptic plane. An even more important issue is the ambient solar wind state. ENLIL used a more realistic solar wind background, whereas DBEM assumed average solar wind conditions for all CMEs (\(w \sim 350\) km s\(^{-1}\), \(\gamma \sim 0.1 \cdot 10^{-7}\) km\(^{-1}\)), not taking into account possible event–to–event variability (e.g., propagation through high-speed streams). Another aspect that was not considered is a possible preconditioning effect, when an earlier CME “clears the path,” effectively reducing \(\gamma\) (Liu et al. 2014), which can be even 10 times lower than the present average value in extreme cases (Temmer & Nitta 2015). Therefore, it is reasonable to assume that DBEM would perform better if a more realistic solar wind background had been considered and if the radial speed in the ecliptic plane were used. Nevertheless, overall these errors are
Table 1

Observed and Calculated Transit Times for the 25 CMEs under Study Using DBEM and ENLIL

| CME Start Date | ICME Arrival Date and Time | Observed TT (hr) | Observed v (km s⁻¹) | DBEM v̂ (km s⁻¹) | v̂ err (km s⁻¹) | DBEM ΔTT (hr) | DBEM v̂ max (km s⁻¹) | DBEM v̂ min (km s⁻¹) | DBEM pᵢ (%) | ENLIL TT (hr) | ΔTT (hr) | ENLIL v̂ max (km s⁻¹) | ENLIL v̂ min (km s⁻¹) | ENLIL pᵢ (%) | ENLIL ΔTT (hr) | ENLIL v̂ max (km s⁻¹) | ENLIL v̂ min (km s⁻¹) |
|----------------|---------------------------|------------------|---------------------|-----------------|--------------|----------------|-------------------|-----------------|---------------|----------------|----------|-------------------|------------------|---------------|----------------|-------------------|------------------|
| 2013           |                           |                  |                     |                 |              |                |                   |                 |               |                 |          |                   |                   |               |                 |                   |                   |
| Apr 11         | Apr 13 22:13              | 62.8             | 520                 | 100             | 48.2         | −14.6         | 40.4              | 56.3            | 653            | 133           | 556     | 807               | 100                | 46.8          | −16              | 41.4               | 52.9              |
| Jun 21         | Jun 23 03:51              | 48.7             | 720                 | 97.9            | 34.8         | −13.9         | 28.5              | 42              | 841            | 121           | 676     | 1137              | 97.9                | 33.9          | −14.8            | 30.3               | 39.6              |
| Aug 02         | ...                       | ...              | 0                   | ...             | CR           | ...            | ...               | ...             | ...            | ...           | 0       | ...               | ...               | ...           | ...             | ...               | ...               |
| Aug 07         | ...                       | ...              | 100                 | 74              | FA            | ...            | ...               | ...             | ...            | ...           | 100     | ...               | ...               | ...           | ...             | ...               | ...               |
| Aug 30         | Sep 02 01:56              | 71.1             | 510                 | 91.7            | 63.3         | −7.8          | 54.5              | 69.4            | 547            | 37            | 488     | 628               | 95.8                | 53.7          | −17.4            | 48.3               | 58.3              |
| Sep 19         | ...                       | ...              | 0                   | ...             | CR           | ...            | ...               | ...             | ...            | ...           | 100     | ...               | ...               | ...           | ...             | ...               | ...               |
| Sep 29         | Oct 02 01:15              | 52.6             | 630                 | 100             | 50.8         | −1.8          | 39.7              | 62.4            | 627            | −3            | 526     | 793               | 100                | 55.5          | 2.9              | 44.7               | 64.6              |
| Oct 06         | Oct 08 19:40              | 53               | 650                 | 100             | 58.3         | 5.3           | 47.5              | 64.6            | 569            | −81           | 503     | 681               | 91.7                | 79.5          | 26.5             | 69.7               | 89.6              |
| Oct 22         | ...                       | ...              | 100                 | 58              | FA            | ...            | ...               | ...             | ...            | ...           | 100     | ...               | ...               | ...           | ...             | ...               | ...               |
| 2014           |                           |                  |                     |                 |              |                |                   |                 |               |               |         |                   |                   |               |                 |                   |                   |
| Jan 07         | Jan 09 19:39              | 49.3             | 480                 | 100             | 26.3         | −23           | 19.8              | 33.9            | 1019           | 539           | 772     | 1474              | 100                | 29.9          | −19.4            | 23                 | 39.1              |
| Jan 30         | Feb 02 23:20              | 78.9             | 470                 | 87.5            | 59           | −19.9         | 44.6              | 78.1            | 563            | 93            | 459     | 721               | 54.2                | 65.7          | −13.2            | 53.4               | 77.6              |
| Jan 31         | ...                       | ...              | 100                 | 68.7           | FA            | ...            | ...               | ...             | ...            | ...           | 100     | ...               | ...               | ...           | ...             | ...               | ...               |
| Feb 12         | Feb 15 12:46              | 79.1             | 450                 | 100             | 58.2         | −20.9         | 49.7              | 72.2            | 562            | 112           | 475     | 663               | 100                | 66.1          | −13              | 57.4               | 79.3              |
| Feb 18         | Feb 20 02:42              | 49.3             | 690                 | 75              | 59.5         | 10.2          | 30.3              | 86.2            | 558            | −132          | 424     | 926               | 80.6                | 63.1          | 13.8             | 34.4               | 80.8              |
| Feb 19         | Feb 23 06:09              | 86.2             | 510                 | 100             | 53.1         | −33.1         | 43.5              | 62.2            | 604            | 94            | 518     | 743               | 88.9                | 68.4          | −17.8            | 55.8               | 81.6              |
| Feb 25         | Feb 27 15:50              | 62.7             | 500                 | 54.2            | 38.2         | −24.5         | 29.5              | 55.2            | 763            | 263           | 580     | 1038              | 83.3                | 45.1          | −17.6            | 33.8               | 65.8              |
| Mar 23         | Mar 25 19:10              | 63.4             | 520                 | 43.8            | 72           | 8.6           | 58.1              | 85.5            | 497            | −23           | 431     | 592               | 79.2                | 69.2          | 5.8              | 56.6               | 81                |
| Mar 23         | ...                       | ...              | 8.3                 | 114.4           | CR           | ...            | ...               | ...             | ...            | ...           | 0       | ...               | ...               | ...           | ...             | ...               | ...               |
| Mar 29         | ...                       | ...              | 50                  | 75.2           | FA            | ...            | ...               | ...             | ...            | ...           | 5.6     | ...               | ...               | ...           | ...             | ...               | ...               |
| Apr 02         | Apr 05 10:00              | 68.4             | 500                 | 31.2            | 43.5         | −24.9         | 37                 | 50.4            | 710            | 210           | 598     | 891               | 87.5                | 53.4          | −15              | 43.3               | 59.7              |
| Apr 18         | Apr 20 10:20              | 45.2             | 700                 | 100             | 37.8         | −7.4          | 31.7              | 43.8            | 780            | 80            | 644     | 1010              | 100                | 40            | −5.2             | 36                 | 46.1              |
| Jun 04         | Jun 07 16:12              | 72.4             | 610                 | 88.9            | 79           | 6.6           | 66                 | 110.8           | 460            | −150          | 354     | 539               | 61.1                | 77.1          | 4.7              | 69                 | 83.9              |
| Jun 19         | Jun 22 18:28              | 73.3             | 410                 | 100             | 79.7         | 6.4           | 68                 | 92.3            | 460            | 50            | 403     | 530               | 100                | 71            | −2.3             | 65.6               | 78.6              |
| Jun 30         | ...                       | ...              | 0                   | ...             | CR           | ...            | ...               | ...             | ...            | ...           | 0       | ...               | ...               | ...           | ...             | ...               | ...               |

Note. pᵢ is the probability of arrival, TT is the transit time, and v is the arrival speed. ΔTᵣᵢ = |TT − TT₀| and Δvᵢ = |v − v₀| are prediction errors, whereas TTᵢ/TTᵢ(max) and vᵢ/₁/₁(max) define the arrival spread (the 95% confidence interval) for TT and v, respectively. CR and FA denote correct rejection and false alarm, respectively (for explanations see the main text).
Table 2: Contingency Table Results, Evaluation Measures, and Prediction Errors for DBEM and ENLIL (Calculated Based on the Sample Presented in Table 1)

|                         | DBEM | ENLIL |
|-------------------------|------|-------|
| (a) Contingency table results |     |       |
| Number of hits          | $a$  | 16    |
| Number of misses        | $c$  | 0     |
| Number of false alarms  | $b$  | 4     |
| Number of correct rejections | $d$ | 5     |
| Number of events        | $N = a + b + d$ | 25   |

(b) Evaluation measures

|                                |       |       |
|--------------------------------|-------|-------|
| Correct rejection rate         | $d/(b + d)$ | 55.6% | 66.7% |
| False alarm rate               | $b/(b + d)$ | 44.4% | 33.3% |
| Correct alarm ratio            | $a/(a + b)$ | 80.0% | 84.2% |
| False alarm ratio              | $b/(a + b)$ | 20.0% | 15.8% |
| Brier score                    | BS (see Equation (6)) | 0.17 | 0.18 |

(c) Prediction errors for $TT$ (hr)

|                                |       |
|--------------------------------|-------|
| Mean error (ME)                | $-9.7$ |
| Mean absolute error (MAE)      | 14.3  |
| root mean square error (RMSE)  | 16.7  |

(d) Prediction errors for $v$ (km s$^{-1}$)

|                                |       |
|--------------------------------|-------|
| Mean error (ME)                | 84    |
| Mean absolute error (MAE)      | 133   |
| root mean square error (RMSE)  | 181   |

would seem that both models slightly underforecast (ENLIL being closer to the line of perfect reliability). Similar conclusions were drawn for ENLIL by Mays et al. (2015) with an extended sample and a slightly different selection of bins (see Figure 9(a) in Mays et al. 2015).

In Figure 5(b) a rank diagram (also known as the “Talagrand” diagram) is shown, which reflects how well the ensemble spread of the forecast represents the true variability of the observations, i.e., whether observations statistically belong to the forecasted distributions (see, e.g., Hamill 2001, and references therein). The rank diagram is constructed by sorting the $n$ ensemble members and the observation for each event from earliest to latest arrival time and “counting” at which place we find the observation with respect to other ensemble members (denoted as “rank”). For a perfect forecast, where we can statistically regard the observation as another member of the ensemble the observation is equally likely to occur in each of the $n + 1$ possible “ranks.” Due to the fact that not all events have the same ensemble sizes and moreover, that ensemble sizes for DBEM are drastically increased by introducing the $w$ and $\gamma$ synthetic values, we follow the same procedure as Mays et al. (2015) and normalize the rank number to $n = 9$ (in total 10 possible ranks). The “perfect reliability” line in Figure 5(b) represents an ideally flat distribution where each rank has the same probability (i.e., 1.6 events per rank).

It can be seen that the rank diagram for both ENLIL and DBEM deviates from the perfect reliability. For both models the rank diagram shows a U-shape, which indicates the undervariability of the ensemble and an asymmetrical shape, which indicates a bias. For both DBEM and ENLIL the possible bias is toward underforecasting, i.e., predicting smaller transit times than observed (also visible from the negative mean error presented in Table 2). This is also supported by the fact that in only 38% of events can the observed $TT$ be found within the DBEM prediction spread, whereas for ENLIL this percentage is somewhat higher (50%). The possible reason for this bias could be related to fast CMEs. Mays et al. (2015) found that in their (extended) sample ENLIL generally predicted fast CMEs to arrive earlier than they were observed.

To confirm this assumption, CME arrival time prediction errors are plotted against the CME input speed in Figure 6. An almost consistent negative prediction error can be seen for both DBEM and ENLIL for fast CMEs above $\sim 1000$ km s$^{-1}$ (cf. Figure 8(a) by Mays et al. 2015). This indicates that fast CMEs are indeed predicted to arrive earlier than observed for both ENLIL and DBEM, and results in a negative mean error and a bias toward underforecasting. This might be related to model limitations, since they do not take into account all relevant physical processes. Liu et al. (2013, 2016) found that fast CMEs (with speeds above $1000$ km s$^{-1}$) differ from slower CMEs by a rapid deceleration process, which is not taken into account by DBEM and ENLIL. This would result in earlier predictions and overestimated arrival speeds. Indeed, as seen in Table 2(d) and Figure 8(a), DBEM has a tendency to overforecast the arrival speed for fast CMEs. Note that the role of CME-driven shocks is not considered in the drag-based model (Reiner et al. 2003; Liu et al. 2013, 2016). DBM considers the physics of the magnetic structure of CMEs, not their related shocks, and it only estimates the shock arrival time using an empirically obtained proxy value of $\gamma$ (Vršnak et al. 2014). Another possible reason could be an overestimation of the CME initial speed for fast CMEs, as suggested by

still comparable to CME arrival time prediction errors reported in other studies (see, e.g., Li et al. 2008; Vršnak et al. 2014; Mays et al. 2015; Sudar et al. 2016; Wold et al. 2017, and references therein).

Next, we test the performance of DBEM using the so-called reliability diagram, which shows how well the predicted probabilities of an event correspond to their observed frequencies, i.e., how well the model predicts the probability of arrival (see, e.g., Jolliffe & Stephenson 2003). In order to obtain the reliability diagram the events were binned according to their forecasted probability of arrival into 5 “bins”: 0%, [0%–33%], [33%–66%], [66%–100%], and 100%. For each bin an observed relative frequency was calculated as $\sum_{i=1}^{N_b} O_i/N_b$, where $N_b$ is the number of events corresponding to the bin and $O_i$ takes a value 0 if event $i$ did not occur and 100 if the event $i$ did occur. The line of perfect reliability is the identity line, i.e., in a perfectly reliable forecast the forecast probability equals the observed relative frequency. The reliability diagram based on the selected sample for DBEM and ENLIL is shown in Figure 5(a), where the number of events used in each calculation is shown next to each point. Although some points are calculated based on only two events being observed, the diagram can reflect some general aspects of the DBEM and ENLIL forecast reliability. It can be seen that for the 100% bin both DBEM and ENLIL overforecast, i.e., predict higher probability of CME arrival than is observed, in agreement with a notable value of the false alarm rate in Table 2. For the 0% bin the point lies on the line of perfect reliability for both ENLIL and DBEM; however, it should be noted that this is related to the fact that there are no missing alarms for either model. With the intermediate bins one should be careful with drawing conclusions, due to the small number of calculations in some points, but at a descriptive level it
Figure 5. (a) Reliability diagram of the forecast probability of CME arrival for the whole sample of 25 events; (b) rank histogram for the 16 hits.

Figure 6. CME arrival time prediction error plotted against the CME input speed for (a) DBEM and (b) ENLIL. The error bars represent the spread in the ensemble forecasts.

Figure 7. (a) ENLIL-calculated vs. DBEM-calculated transit time; (b) observed vs. calculated transit time for ENLIL (blue) and DBEM (red). The solid lines represent the linear best fits, with corresponding equations and correlation coefficients $r$. The dotted line represents the identity line (perfect match). The error bars represent the spread in the ensemble forecasts.
Although DBEM and ENLIL use a range of CME initial speeds as input, a large systematic overestimation of CME speed in the ensemble (e.g., due to a limited number of measurement points in fast events) could lead to the observed underforecast of transit times and overforecast of arrival speed.

Finally, we examine the correlation between observed and calculated transit times. In Figure 7(a) ENLIL-calculated transit time is plotted as a function of DBEM-calculated transit time. It can be seen that there is some scatter and the linear best fit somewhat deviates from the identity line, but in general there is good agreement between ENLIL-calculated and DBEM-calculated transit times, similar to the results obtained for non-probabilistic ENLIL and DBM by Vršnak et al. (2014). In Figure 7(b), the observed transit time is plotted as a function of calculated transit time for ENLIL (blue) and DBEM (red). It can be seen that both for ENLIL and DBEM the linear best fit substantially deviates from the identity line. This is seen in the linear best-fit coefficients, where the slope is much smaller than 1 and the intercept is much larger than 0. Furthermore, it can be seen that for higher calculated transit times the scatter is around the identity line, whereas for smaller calculated transit times the scatter is above the identity line, which is related to the fact that fast CMEs are predicted to arrive earlier than observed. In Figure 8(b), the observed arrival speed is plotted as a function of calculated arrival speed for DBEM. Similar to transit times, it can be seen that the linear best fit substantially deviates from the identity line, related to the overforecast of the arrival speed (for smaller values of calculated arrival speeds the scatter is around the identity line, whereas for larger values of calculated arrival speeds the scatter is below the identity line). We note that there seems to be an outlier with calculated \( v = 1019 \text{ km s}^{-1} \). When the outlier is removed the slope of the best linear fit obtains a value of 0.36 and thus improves toward the identity line, and the correlation coefficient increases to \( r = 0.42 \). We note that the outlier seems not to appear in the transit time calculation (Figure 7(b)), but does appear in the calculation of arrival speed (Figure 8(b)).

**4. Summary and Conclusion**

We present a probabilistic approach to drag-based CME propagation modeling, the drag-based ensemble model (DBEM), which is available as an online tool at Hvar Observatory website,\(^8\) and as one of the European Space Agency (ESA) space situational awareness (SSA) products.\(^9\) An ensemble of \( n \) members is used as a CME input, where the variation in model parameters is also taken into account by constructing the synthetic values of solar wind speed, \( w \), and drag parameter, \( \gamma \), \( m \) synthetic values of \( w \) and \( \gamma \) are constructed under the assumption that the measurements are behaving as random variables and follow a normal distribution. The model input thus consists of \( n \cdot m^2 \) ensemble members, where \( n \) is restricted by the observations and the optimal \( m \) was found as a compromise between the convergence of the output distribution and the computing time. The outputs are the probability of arrival and distributions of CME transit times and arrival speeds, where the median is taken as the most likely value and the range is defined with a 95% confidence interval. Therefore, DBEM is able to provide probabilities of CME arrival at a specific target together with the uncertainties in the arrival time and speed.

We evaluated the performance of the new DBEM model on a refined sample taken from the study by Mays et al. (2015), where it can also be compared to the performance of ENLIL. Depending on the number of CPUs and their speed, DBEM can perform a thousand (on a single CPU) or more runs per second and is several orders of magnitude faster than ENLIL, which is its main advantage. Comparison between DBEM and ENLIL revealed that ENLIL performs slightly better than DBEM, most probably owing to the more realistic background solar wind conditions and a more suitable CME input (3D velocity). Despite small differences, the performance of the two models is similar and comparable to a “standard” CME prediction error in transit time of \( \sim 10 \text{ hr} \). We find that for this particular sample, in both models fast CMEs are predicted to arrive earlier than observed, which is related to model limitations and possibly also to the overestimation of the CME initial speed for fast CMEs. Since both models show similarly good, as well as bad

---

\(^8\) http://oh.geof.unizg.hr

\(^9\) http://swe.ssa.esa.int/heliospheric-weather.
results, additional actions should probably be taken in order to improve the overall CME arrival forecast. A possible improvement might be to use different propagation models for different solar activity conditions. This approach, as well as the ensemble forecasting, would be analogous to the methods used in meteorology, where different models are used for different conditions, since there is not a single model that is capable of forecasting the weather for all types of regions. In line with this view, the “CME Scoreboard” website\textsuperscript{10} can serve as the platform to compare different models simulating forecasting a variety of CME events occurring in real-time. Anyone is invited to submit their estimate of the arrival time of a recently observed CME in real-time to the CME Scoreboard. Anyone is invited to submit their estimate of the arrival time of a recently observed CME in real-time to the CME Scoreboard. The CME Scoreboard can be used as the platform to compare different models simulating the ensemble forecast from a variety of models and methods. The possible benefits from this type of approach are yet to be evaluated in future studies.

The research leading to these results has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No. 745782. B.V., J.Č., and M.D. acknowledge financial support by the Croatian Science Foundation under project 6212 “Solar and Stellar Variability.” M.T. acknowledges the support by the FFG/ASAP Programme under grant No. 859729 (SWAMI). A.M.V. and M.T. acknowledge support from the Austrian Science Fund (FWF): P24092-N16 and V195-N16. This work was supported by the Scientific & Technological Cooperation between Austria-Croatia under grant No. HR 24/2018.

\textbf{ORCID iDs}

Mateja Dumbović \texttt{https://orcid.org/0000-0002-8680-8267}
Jaša Čalogović \texttt{https://orcid.org/0000-0002-4066-726X}
Bojan Vršnak \texttt{https://orcid.org/0000-0002-0248-4681}
Manuela Temmer \texttt{https://orcid.org/0000-0003-4867-7558}
M. Leila Mays \texttt{https://orcid.org/0000-0001-9177-8405}

\textbf{Astrid Veronig} \texttt{https://orcid.org/0000-0003-2073-002X}
Isabell Piantschitsch \texttt{https://orcid.org/0000-0001-9124-6644}

\textbf{References}

Cargill, P. J., Chen, J., Spicer, D. S., & Zalesak, S. T. 1996, \textit{JGR}, 101, 4855
Crosby, N. B., Veronig, A., Robbrecht, E., et al. 2012, in AIP Conf. Proc. 1500, Space Weather: The Space Radiation Environment: 11th Annual Int. Astrophysics Conf., ed. Q. Hu et al. (Melville, NY: AIP), 159
Dryer, M., Fry, C. D., Sun, W., et al. 2001, \textit{SoPh}, 204, 265
Dumbović, M., Srivastava, N., Rao, Y. K., et al. 2017, \textit{SoPh}, 292, 96
Gopalswamy, N., Lara, A., Lepping, R. P., et al. 2000, \textit{GeoRL}, 27, 145
Gopalswamy, N., Lara, A., Yashiro, S., Kaiser, M. L., & Howard, R. A. 2001, \textit{JGR}, 106, 29207
Hamill, T. M. 2001, \textit{MWRv}, 129, 550
Hess, P., & Zhang, J. 2014, \textit{ApJ}, 792, 49
Jolliffe, I. T., & Stephenson, D. B. 2003, Forecast Verification: A Practioner's Guide in Atmospheric Science (New York: Wiley)
Li, H. J., Wei, F. S., Feng, X. S., & Xie, Y. Q. 2008, \textit{JGR}, 113, A09101
Liu, Y. D., Hu, H., Wang, C., et al. 2016, \textit{ApJS}, 222, 23
Liu, Y. D., Luhmann, J. G., Kajdić, P., et al. 2014, \textit{NatCo}, 5, 3481
Liu, Y. D., Luhmann, J. G., Laguz, N., et al. 2013, \textit{ApJ}, 769, 45
Mays, M. L., Takakшивили, A., Pulkkinen, A., et al. 2015, \textit{SoPh}, 290, 1775
McComas, D. J., Angold, N., Elliott, H. A., et al. 2013, \textit{ApJ}, 779, 2
Odstrcil, D., Riley, P., & Zhao, X. P. 2004, \textit{JGR}, 109, A02116
Reiner, M. J., Kaiser, M. L., & Bougeret, J.-L. 2003, in AIP Conf. Ser. 679, Solar Wind Ten, ed. M. Velli et al. (Melville, NY: AIP), 152
Rollett, T., Möstl, C., Isavnin, A., et al. 2016, \textit{ApJ}, 824, 131
Sachdeva, N., Subramanian, P., Colaninno, R., & Vourlidas, A. 2015, \textit{ApJ}, 809, 158
Sachdeva, N., Subramanian, P., Vourlidas, A., & Bothmer, V. 2017, \textit{SoPh}, 292, 118
Sheeley, N. R., Walters, J. H., Wang, Y.-M., & Howard, R. A. 1999, \textit{JGR}, 104, 24739
Sudar, D., Vršnak, B., & Dumbović, M. 2016, \textit{MNRAS}, 456, 1542
Takahashi, T., & Shibata, K. 2017, \textit{ApJL}, 837, L17
Temmer, M., & Nitta, N. V. 2015, \textit{SoPh}, 290, 919
Temmer, M., Rollett, T., Möstl, C., et al. 2011, \textit{ApJ}, 743, 101
Temmer, M., Vršnak, B., Rollett, T., et al. 2012, \textit{ApJ}, 749, 57
Vršnak, B., Temmer, M., Žic, T., et al. 2014, \textit{ApJS}, 213, 21
Vršnak, B., & Žic, T. 2007, \textit{A&A}, 472, 937
Vršnak, B., Žic, T., Vrbanec, D., et al. 2013, \textit{SoPh}, 285, 295
Wold, A. M., Mays, M. L., Takakшивили, A., et al. 2017, arXiv:1801.07818
Wu, C.-C., Dryer, M., Wu, S. T., et al. 2011, \textit{IGRA}, 116, A12103
Zhao, X., & Dryer, M. 2014, \textit{SpWea}, 12, 448
Zhao, X., Liu, Y. D., Inhester, B., et al. 2016, \textit{ApJ}, 830, 48
Žic, T., Vršnak, B., & Temmer, M. 2015, \textit{ApJS}, 218, 32

\footnotesize{\textsuperscript{10}https://kauai.ccmc.gsfc.nasa.gov/CMEscoreboard/}