1 The general description

The Drag-Based Ensemble Model version 3 (DBEMv3) provides probabilistic predictions in ecliptic plane of the Coronal Mass Ejection (CME) arrival time and speed at Earth or any given target in the solar system (planets and satellites). DBEM calculations are based on the Drag-Based Model (DBM) for heliospheric propagation of CMEs, which is a widely used analytical model to predict CME propagation at a given heliospheric location [for details see Vršnak et al., 2013]. DBM assumes that the propagation of CMEs in interplanetary space is solely under the influence of magnetohydrodynamical (MHD) drag, where CME propagation is determined based on CME initial properties as well as the properties of the ambient solar wind [Vršnak, 2001, Vršnak and Žic, 2007, Vršnak et al., 2010]. The assumed CME 2D geometry for DBM is a cone and the CME leading edge is initially a semicircle defined by the CME angular width that flattens with time [Žic et al., 2015, Dumbović et al., 2021]. The model also assumes a constant solar wind speed, $w$, and drag parameter, $\gamma$, which is in general valid for distances beyond 15 solar radii ($R > 15 R_\odot$). In such conditions 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 at the CME expansion [for details see e.g. Žic et al., 2015, and references therein].
The probabilistic aspect is introduced using ensemble modelling as described by Čalogović et al. [2021]. The user provides observational input values and uncertainties for selected input parameters from which the DBEMv3 generates an ensemble. The ensemble is produced using random values that follow a normal distribution with observational input value as mean and standard deviation derived from uncertainty (uncertainty = 3σ). The output is generated through multiple runs of DBM for each ensemble member, where typically 10,000 or more runs are needed to get the reliable results (i.e. output distributions). The output consists of the calculated probability of arrival, most likely CME arrival time and arrival speed, statistical information about output distributions as well as overview of the input ensemble (detailed description is given in the following sections).

Note that DBM/DBEM describes the physics of CME magnetic structure propagation and not of the corresponding shock. However, the DBM-ENLIL comparison indicates that although the DBM describes CME-ejecta propagation, it could also be applied as a proxy of the CME-shock propagation using lower values of the drag parameter, typically \( \gamma = 0.1 \times 10^{-7}\text{km}^{-1} \) [Vršnak et al., 2014, Dumbović et al., 2018]. Detailed description of DBEMv3 and comparison with previous DBEM versions as well as validation and evaluation of model is presented in Čalogović et al. [2021].

Since DBEMv3 version, the DBM tool is integrated in the same package allowing user to run the simple DBM calculation for a given set of input parameters without employing the uncertainties and producing the ensembles. DBM also shows the CME propagation visualizations such as CME geometry and kinematics (see example in Figure 1). In DBEMv3 is also implemented Graduated Cylindrical Shell (GCS) model option [Thernisien et al., 2006, 2009, Thernisien, 2011] to represent the flux rope structure of CMEs. GCS assumes that geometrically a CME can be described as a hollow croissant consisting of a tubular section forming the main body of the structure which is attached to two cones that correspond to CME “legs”. It provides an improved cross-section derivation for calculating more reliably the CME propagation in the ecliptic plane and it was recently chosen as the recommended model for CME measurements [Verbeke et al., 2019]. GCS model option allows user to calculate CME’s angular half-width (\( \lambda \)) from GCS \( \alpha, \kappa \) and tilt parameters [for details see Čalogović et al., 2021].

2 Input description

DBEMv3 input and calculations consist of two steps. In the first step user provides the standard DBM input parameters needed to run a single DBM calculation and show DBM results together with CME geometry and kinematics visualizations. In the second step user defines uncertainties for each input parameter that are needed to run DBEM and to calculate the probability of arrival as well as CME transit time and arrival speed probability distributions.
Figure 1: Example of DBM visualization implemented in DBEMv3 for CME launched on 30 August 2013 at 6:21 UTC.  
a) CME geometry plot where the CME is marked with red shaded area and CME apex with a red dashed line. 
b) CME kinematic plot with CME distance, $R$ (blue solid line), CME speed, $v$ (red solid line) and CME acceleration, $a$ (green dashed line) plotted against the time (in UTC) where the first point on x-axis corresponds to CME launch time and the last point to time when CME hits the target (Earth).
2.1 DBM input

In the first step (on the first page) user is required to define all basic DBM input parameters ($x_i$) that consist of initial CME properties: date and time, radial distance ($R_0$), start speed ($v_0$), angular half-width ($\lambda$) and longitude ($\phi_{CME}$), and model parameters: solar wind speed ($w$), drag parameter ($\gamma$) and target. All input parameters are given and shortly described in Table 1.

The initial CME properties are determined by the user using an arbitrary measurement method ranging from plane-of-sky measurements to 3D reconstruction methods with the input quite similar to other CME heliospheric propagation models [e.g. Vršnak et al., 2014, Dumbović et al., 2018, and references therein] and available through services such as e.g. ESA SSA S-ESC tools. In the case that user employs Graduated Cylindrical Shell (GCS) model for CME reconstruction, DBEMv3 gives the option to calculate CME’s angular half-width, $\lambda$ from GCS $\alpha$ (angular half-with between the legs), $\kappa$ (CME aspect ratio) and tilt (angle) parameters [for details see Thernisien, 2011, Dumbović et al., 2019, Čalogović et al., 2021].

The model parameters are also typically determined by the user, however, the user can select some of the proposed values for drag parameter $\gamma$ and solar wind speed, $w$. The proposed values of $\gamma$ are dependent on the type of the event (slow, normal, fast) and are empirical-based [Vršnak et al., 2013, 2014, Žic et al., 2015]. Solar wind speed values are provided from the Empirical Solar Wind Forecasting (ESWF) tool available at ESA SSA H-ESC, which makes a 4 days of solar wind speed estimation at Earth based on the observed areas of coronal holes in a narrow heliographic longitudinal slit [for further details see Vršnak et al., 2007, Rotter et al., 2012, Reiss et al., 2016].

After the user enters all input parameters ($x_i$ values) at the input page in the first step and submits them by clicking on "Run DBM and set DBEM uncertainties", DBEMv3 performs a simple DBM calculation and shows DBM results with the CME geometry and kinematic plots (Figure 1, see further description in Section 3.1) and provides also the new input page with uncertainties.

2.2 Uncertainties and DBEM input

After running the DBM calculation, user can access the uncertainties page by clicking on "Proceed to DBEM uncertainties" where it is possible to select a numerical value for the uncertainty of the each input parameter (date and time, $R_0$, $v_0$, $\lambda$, $\phi_{CME}$, $w$ and $\gamma$). If the uncertainty of some input parameter is set to 0, DBEMv3 will only use input parameter without uncertainty. The uncertainties cannot be arbitrarily large. The maximum allowed uncertainty for CME time and longitude is constrained based on "common sense" and amounts to 1440 minutes (24 hours) and 120 degrees, respectively. For other parameters the uncertainty value is constrained by the range of possible values for the corresponding
Table 1: Input parameters for DBEMv3.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CME date (at $R_0$)</td>
<td>the date when the CME apex is at distance $R_0$; select from the dropdown menu</td>
</tr>
<tr>
<td>CME time in UTC (at $R_0$)</td>
<td>the time when the CME apex is at distance $R_0$; select from the dropdown menu</td>
</tr>
</tbody>
</table>
| Drag parameter, $\gamma$        | drag-parameter in $\times 10^{-7}$ km$^{-1}$; select from the dropdown menu (based on CME speed):  
|                                 | 0.5 for “slow” CMEs ($v < 500$ km/s)  
|                                 | 0.2 for “normal” CMEs ($500$ km/s < $v < 1000$ km/s)  
|                                 | 0.1 for “fast” CMEs ($v > 1000$ km/s)  
|                                 | or can be entered manually ($0.001 \leq \gamma \leq 100$) |
| Solar wind speed, $w$            | the background constant solar wind speed in km/s; entered manually ($200$ km/s < $w$ < $800$ km/s); a solar wind forecast value derived from ESWF tool is provided next to input box (only valid for Earth!) |
| CME starting radial dist., $R_0$ | distance of the CME apex for a given date & time (in $R_{\text{SUN}}$) entered manually ($1R_{\text{SUN}} < R_0 < 250R_{\text{SUN}}$) |
| Starting speed of CME, $v_0$     | the speed of CME apex at $R_0$ in km/s; entered manually ($50$ km/s < $v$ < $5000$ km/s) |
| CME’s angular half-width, $\lambda$ | The angular half-width of the cone which represents the geometry of the CME (in degrees) entered manually ($0$ deg < $\lambda$ < $90$ deg, HALO CME $\lambda = 89$ deg) |
| GCS input button allows to calculate $\lambda$ from GCS $\alpha$, $\kappa$ and tilt parameters |
| Long. of CME source reg., $\varphi_{\text{CME}}$ | the longitudinal position of the CME apex for a given date & time in HEE system (in degrees); entered manually ($-180$ deg < $\varphi_{\text{CME}}$ < $180$ deg) |
| Target                           | planet/spacecraft for which the calculation will be performed; select from the dropdown menu |
| Number of DBEM runs, $n$         | the number of ensemble members can be entered manually ($1000 \leq n \leq 100 000$) or selected from the dropdown menu (based on precision):  
|                                 | 10 000 for fast preview  
|                                 | 30 000 for normal precision  
|                                 | 50 000 for more accurate results  
| Note: larger number of runs will need more time to calculate |
parameter (given in Table 1). Note that uncertainty of the starting radial distance cannot be defined, since it is implicitly included in the uncertainty of the start time. In addition, it is not possible to define the uncertainty of the target (or target position) because this value is extracted from the ephemeris database embedded in the tool. Namely, DBEMv3 uses for each possible target (e.g. planets and satellites) predefined ephemeris data\(^2\) and it takes also in account movement of targets during CME transit time.

DBEMv3 produces ensembles under the assumption that the real measurements of input parameters \(X_i\) follow a normal distribution and can be defined in the form \(X_i = \bar{x}_i \pm \Delta x_i\), where \(\bar{x}_i = \mu_i\) is the mean of the normal distribution and \(\Delta x = 3\sigma\) defines a range where 99.7% of measurements are found. In this case DBEMv3 input parameters are \(\bar{x}_i\) and uncertainties are defined as \(3\sigma\). Thus, for each input parameter with uncertainty, random samples are drawn from normal distribution determined by \(\mu_i\) and \(\sigma\) where the number of samples correspond to the number of the ensemble members and therefore the number of DBEM runs (can be selected by the user). The samples determine the probability density functions, PDFs for each parameter, which are limited to \(3\sigma\) range and are given in the output page after DBEMv3 calculation is performed. Example of PDFs for all input parameters is shown in Figure 2.

At the bottom of the uncertainties page, the user chooses the number of DBEMv3 runs, \(n\) which determines the number of ensemble members. Due to its probabilistic nature the number of DBEMv3 runs is constrained with a minimum of 1,000 runs. On the other hand, due to the calculation-time the number of DBEMv3 runs is constrained with a maximum of 100,000 runs. Input parameter PDFs depend on \(n\) as they are generated as random samples drawn from the normal distribution. Therefore, if the number of ensemble members is limited, every DBEM calculation using the same input can produce slightly different results, with deviations decreasing as \(n\) is increasing. This is shown in Figure 3 where the deviations in calculated hit probability (Figure 3a), transit time (Figure 3b) and arrival speed (Figure 3c) given as \(2\sigma\) (95\% confidence interval) are plotted against the number of DBEM runs. For all three output parameters very similar exponential dependences were found. In all cases for 20,000 DBEM runs deviations almost converge to the constant value (less than 5 minutes for transit time, 0.5 km/s for arrival speed and 0.1\% for hit probability) what is several orders of magnitude smaller than the typical errors due to lack of reliable input data or DBM model constrains (e.g. CME-CME interactions). Thus, it is our recommendation to use \(n \geq 20,000\) for reliable results. Since larger number of runs increases the calculation time there is also an option of fast preview (\(n = 10,000\)). Although the allowed maximum is 100,000 runs, we advise the user not to use it except in very extreme cases (due to longer calculation time). For most purposes 30,000 DBEMv3 runs will give very accurate results and larger number of runs won’t significantly lower (already very small) deviations due to the random sample method.

Please note that on input and uncertainties pages a "Reset" button is provided to return

\(^2\) taken from JPL’s HORIZONS system website https://ssd.jpl.nasa.gov/horizons.cgi
to the default values. However, note that once the user proceeds to the uncertainty page it is not possible to reset also input parameters by using "Reset" button on the uncertainties page but rather "Reset" button should be used on the input page or the whole page should be reloaded.

3 Output description

3.1 DBM output

When the user submits all input parameters on the first page (described in Section 2.1), DBEMv3 runs DBM module and shows the page with the DBM results (without the uncertainties). This includes CME arrival date and time, transit time \((TT)\) in hours, CME speed and acceleration at target. Calculations are performed using 2D flattening DBM, a 2D model which combines basic DBM and cone geometry describing the propagation of the CME leading edge which does not evolve self-similarly [for details see Dumbović et al., 2021] that is also the same model employed later in DBEM calculations. Beside the basic DBM output and all input parameters listed, CME geometry animation and CME kinematic plot are also shown on the page with DBM results (Figure 1). Depending on the target, CME geometry plots shows all planets (depicted as circles) and satellites (depicted as triangles) at the time of CME where CME itself is marked with red shaded region and CME apex as red dotted line (Figure 1a). CME kinematic plot is produced for CME distance, speed and acceleration versus time on x-axis from the CME launch time until the time CME hits the target (Figure 1b). The user is also given an option to download all DBM results as zip file.

3.2 DBEM output

After entering all needed uncertainties DBEMv3 performs calculations and produces the output on a separate "DBEM results" page. On the top of the page DBEMv3 shows the probability of CME arrival at target, most probable CME arrival time and CME speed at target displayed in textual form. This is followed by a Figure with a pie chart plot of the probability of arrival (hit target statistics) and histograms of the transit time and arrival speed distributions (see example in Figure 4). The DBEM statistics table with the main distribution parameters (median, mean, confidence intervals, standard deviation, min, max) for transit time, arrival speed and acceleration is also provided. Finally, Figure for the DBEM input is shown in the form of histograms that give the distributions of all input parameters for the whole ensemble (Figure 2). All DBEMv3 output parameters are listed in Table 2.

For each ensemble member, \(i.e.\) individual run, DBEMv3 calculates whether the CME will hit or miss the target (\(e.g.\) Earth) and for the whole ensemble the probability of the arrival is calculated as \(p = n_{\text{hits}}/n_{\text{tot}},\) where \(n_{\text{hits}}\) is the number of ensemble members that
Figure 2: Example of the six input parameters with their corresponding uncertainties for CME launched on 30 August 2013 at 6:21 UTC shown in histograms as probability density functions (PDFs). Given uncertainties (99.7% confidence intervals or 3σ) are indicated above each plot (lower uncertainty < input value < upper uncertainty). The blue bars represent the DBEM samples (in this case 30 000 for each parameter) and the red solid line denotes the calculated normal distribution.
Figure 3: Dependence of standard deviations in DBEM results due to randomly generated input probability density functions on the number of performed DBEM runs calculated for: (a) probability that CME hits the target, $p_{\text{tar}}$ (%), (b) CME transit time, $T_T$ (minutes) and (c) CME arrival speed at target, $v_{\text{tar}}$ (km s$^{-1}$). Red ($p_{\text{tar}}$), green ($T_T$) and blue ($v_{\text{tar}}$) points represent standard deviations given as $2\sigma$ (95% confidence interval) and calculated by running 20 times DBEM with the identical input data but the different number of runs. The best power law fit is denoted with the black dotted curve.
Figure 4: Example of DBEM results for CME launched on 18 February 2014 at 5:20 UTC. The upper left panel shows table with all input parameters and their uncertainties. The upper right panel shows pie chart with the hit target statistics (miss or hit). Lower panels show histograms with transit time (left) and CME speed at target, $v_{\text{tar}}$ (right). Median value is denoted on histograms as black solid line and mean value as blue dashed line. Calculated 95% confidence intervals are shown by red dotted lines.
Table 2: Output parameters for DBEMv3 (*CI=confidence interval).

<table>
<thead>
<tr>
<th>Output</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of CME arrival</td>
<td>calculated as $p = \frac{n_{\text{hits}}}{n_{\text{tot}}}$ (in %) displayed as text and Pie Chart (red and blue color for hits and misses, respectively)</td>
</tr>
<tr>
<td>CME arrival time</td>
<td>date &amp; time of the CME arrival at target (in UTC) text display of median (most likely) value and the 95% CI*</td>
</tr>
<tr>
<td>CME speed, $v_{\text{tar}}$</td>
<td>CME arrival speed at target (in km/s) the results are displayed as: text display of median/most likely value and the 95% CI* $v_{\text{tar}}$ distribution/histogram plot for hits statistical properties table of $v_{\text{tar}}$ distribution for hits</td>
</tr>
<tr>
<td>CME Transit Time, $TT$</td>
<td>transit time from $R_0$ to target distance (in hours) the results are displayed as: $TT$ distribution/histogram plot for hits statistical properties table of $TT$ distribution for hits</td>
</tr>
<tr>
<td>CME acceleration, $a_{\text{tar}}$</td>
<td>CME arrival acceleration at target (in km/s$^2$) statistical properties table of acceleration distribution for hits</td>
</tr>
<tr>
<td>DBEM input plots</td>
<td>histograms of distributions of input parameters relative frequency vs corresponding input parameter (CME date &amp; time, $\gamma$, $w$, $v_0$, $\lambda$, and $\varphi_{\text{CME}}$)</td>
</tr>
</tbody>
</table>

are calculated to hit target and $n_{\text{tot}}$ is the total number of all ensemble members [Dumbović et al., 2018, Čalogović et al., 2021].

The most likely transit time, $TT$ and arrival speed at target, $v_{\text{tar}}$ are determined based on the corresponding distributions, calculated only based on the runs for ensemble members calculated to hit target (the red part of the pie chart in upper right panel of Figure 4). The expected range is given by the 95% confidence interval (CI) and the median is taken as the most likely value. Therefore, CME arrival time, CME transit time and CME arrival speed are displayed in the following textual form: lower 95% CI < median < upper 95% CI. As shown in Figure 4 example, the corresponding distributions for $TT$ and $v_{\text{tar}}$ are presented in histogram plots (the relative frequency is given on the y-axis) together with some main distribution parameters (mean, median and 95% CI). $TT$ and $v_{\text{tar}}$ distribution parameters are also displayed in the table "DBEM statistics", which also provides distribution parameters for CME arrival acceleration, $a_{\text{tar}}$. If CME completely misses the target (0% chance to hit the target), DBEM won’t show CME arrival time and speed as well as the corresponding histogram plots and statistics.

DBEM input parameters (see example in Figure 2) are also displayed as blue bars in histogram plots (with relative frequency at the y-axis) where given uncertainties (99.7% or
The red solid lines represent calculated normal distributions. The purpose of these DBEM input distribution plots is to serve as a check-up of the input ensemble for the user. Near the bottom of the "Results" page performance information is given as well as instructions where the results have been stored and how to access them. All files with DBEM results (figures, statistics, input parameters, calculation summary and log file) can be downloaded as zip file. Note that each DBEM run is stored on the server for 100 days and it is accessible within that time period.

References


