

Space Weather

RESEARCH ARTICLE

10.1029/2018SW002046

Special Section:

Space Weather Capabilities Assessment

Key Points:

- We have fully defined metadata for CME arrival and impact and showed it is critical for assessment
- The validation event set will not be based on a handpicked set of events, but on two selected time periods
- We have identified metrics for CME arrival time and impact by community consensus

Correspondence to:

C. Verbeke,
christine.verbeke@kuleuven.be

Citation:

Verbeke, C., Mays, M. L., 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, 6–26. <https://doi.org/10.1029/2018SW002046>

Received 31 JUL 2018

Accepted 14 NOV 2018

Accepted article online 29 NOV 2018

Published online 3 JAN 2019

Benchmarking CME Arrival Time and Impact: Progress on Metadata, Metrics, and Events

C. Verbeke¹, M. L. Mays², M. Temmer³, S. Bingham⁴, R. Steenburgh⁵, M. Dumbović^{3,6}, M. Núñez⁷, L. K. Jian², P. Hess⁸, C. Wiegand², A. Taktakishvili^{2,9}, and J. Andries^{1,10}

¹Centre of Mathematical Plasma-Astrophysics, Departement of Mathematics, Catholic University of Leuven, Leuven, Belgium, ²NASA Goddard Space Flight Center, Greenbelt, MD, USA, ³Institute of Physics, University of Graz, Graz, Austria, ⁴Met Office, Exeter, UK, ⁵NOAA/Space Weather Prediction Center, Boulder, CO, USA, ⁶Hvar Observatory, Faculty of Geodesy, University of Zagreb, Zagreb, Croatia, ⁷Department of Languages and Computer Sciences, Universidad de Málaga, Málaga, Spain, ⁸Space Science Division, Naval Research Laboratory, Washington, DC, USA, ⁹Catholic University of America, Washington, DC, USA, ¹⁰Solar-Terrestrial Center of Excellence, Royal Observatory of Belgium, Brussels, Belgium

Abstract Accurate forecasting of the arrival time and subsequent geomagnetic impacts of coronal mass ejections (CMEs) at Earth is an important objective for space weather forecasting agencies. Recently, the CME Arrival and Impact working team has made significant progress toward defining community-agreed metrics and validation methods to assess the current state of CME modeling capabilities. This will allow the community to quantify our current capabilities and track progress in models over time. First, it is crucial that the community focuses on the collection of the necessary metadata for transparency and reproducibility of results. Concerning CME arrival and impact we have identified six different metadata types: 3-D CME measurement, model description, model input, CME (non)arrival observation, model output data, and metrics and validation methods. Second, the working team has also identified a validation time period, where all events within the following two periods will be considered: 1 January 2011 to 31 December 2012 and January 2015 to 31 December 2015. Those two periods amount to a total of about 100 hit events at Earth and a large amount of misses. Considering a time period will remove any bias in selecting events and the event set will represent a sample set that will not be biased by user selection. Lastly, we have defined the basic metrics and skill scores that the CME Arrival and Impact working team will focus on.

1. Introduction

One of the most important phenomena affecting space weather are coronal mass ejections (CMEs; Gosling, 1993; Hudson et al., 2006; Koskinen & Huttunen, 2006). They consist of large-scale eruptions of magnetized plasma, erupting from the Sun, and can occur on a daily basis, especially during solar maxima (Webb & Howard, 1994). CMEs that are Earth directed can directly impact some specific industry sectors, such as space missions, aviation, and electricity networks, but they can also induce indirect effects on for example our navigation systems or gas and oil pipe lines (Schrijver et al., 2015, and references therein). These effects manifest themselves in the form of geomagnetic storms, and a few parameters are of high importance when trying to predict/forecast the severity of the impact of an Earth-directed CME. So far, the space weather community has focused on two main parameters when discussing the performance of CME arrival models: the arrival time of the CME and the arrival speed of the CME (see, e.g., Dumbović et al., 2018; Mays, Taktakishvili, et al., 2015; Vršnak et al., 2014; Wold et al., 2018). Recently, some of the interest has shifted to the prediction of the magnetic B_z component as well, due to the introduction of space weather models capable of inserting flux rope models that include a magnetic field structure (Jin et al., 2017; Kay et al., 2017; Shiota & Kataoka, 2016). The B_z component of the CME, due to the magnetic flux rope that is embedded, is a major driver for the strongest magnetic storms (Huttunen et al., 2005). The progress made so far by the community to predict B_z and to determine appropriate metrics is not the focus of this paper.

When forecasting CME arrivals (both the arrival time and the impact), it is important to keep in mind that some CME propagation models are using 3-D CME input parameters obtained from observations. Those 3-D input parameters on their own contain errors. Typically, coronagraph images from the Large Angle and Spectrometric Coronagraph Experiment (LASCO; Brueckner et al., 1995) instrument on the SOLar and Heliospheric

Observatory (SOHO; Domingo et al., 1995), and the Sun Earth Connection Coronal and Heliospheric Investigation (SECCHI; Howard et al., 2008) instruments on the Solar TERrestrial RELations Observatory (STEREO; Kaiser et al., 2008) Ahead (A)/Behind (B) spacecraft are used. To obtain 3-D CME input parameters, such as the CMEs velocity and angular width, different methods and models have been developed, such as in Zhao et al. (2002), Xie et al. (2004), Xue et al. (2005), Thernisien et al. (2009), Thompson (2009), Millward et al. (2013), Möestl et al. (2014), and Mays, Taktakishvili, et al. (2015). Most of these techniques assume that the CME propagates with a constant width through the corona. So far no systematic studies have been performed that give a better and more accurate overview of the size of the errors associated with the 3-D CME parameters obtained from coronagraph images. Different reconstruction methods can present a rather wide spread in the obtained 3-D parameters on a case by case basis (see Mierla et al., 2010). Once the 3-D CME kinematics parameters are determined from observations, they are typically inserted into the propagation domain of the models as a spherical shape such as in Wang-Sheeley-Arge (WSA)-ENLIL+Cone (Odstrčil et al., 2004) and EUHFORIA (EUropean Heliospheric FORecasting Information Asset, Pomoell & Poedts, 2018). Note that for WSA-ENLIL+Cone, the word Cone refers to the Cone approximation for determining the CME parameters from coronagraph images that can be used as input to the model. Generally, a CME disturbance is inserted at the inner boundary as slices of a homogeneous spherical plasma cloud, however ENLIL also supports other CME shapes such as an ellipsoid. For EUHFORIA, a slightly different method is used as explained in detail by Pomoell and Poedts (2018) and Scolini et al. (2018). Other models such as the drag-based models in Vršnak et al. (2013), Dumbović et al. (2018), and Hess and Zhang (2015) also use observational 3-D CME input parameters.

Generally, most researchers perform their own validation studies (see, e.g., Fry et al., 2001; Gopalswamy et al., 2005; McKenna-Lawlor et al., 2006; Taktakishvili et al., 2009; Vršnak et al., 2014; Mays, Taktakishvili, et al., 2015; Paouris & Mavromichalaki, 2017a; Dumbović et al., 2018; Wold et al., 2018). Most of these validation studies use their own CME event set, CME parameters, and set of metrics. Some effort has been made to compare models such as in Dumbović et al. (2018), where the performance of the Drag-Based Ensemble Model (DBEM) is compared with WSA-ENLIL by using the same set of events. However, when it comes to improving a CME arrival model, as well as the opportunity for users to be able to assess different models, it is best practice to construct general metrics and validation methods, together with a set of CMEs to model and validate against.

From a recent study on the CME arrival time scoreboard (<https://kauai.ccmc.gsfc.nasa.gov/CMEscoreboard/>) (Riley et al., 2018) as well as discussion within the community, it has become apparent that current validation methods can be improved in many ways. First, it is important that the CME event set being considered is a valid representation of the types of CMEs that are observed so that the percentage of hits and misses is very similar to observations and so to operational conditions. Second, to be able to critically evaluate a model's strengths and weaknesses, robust community-agreed metrics are needed. Third, in order to reproduce past validations, and to compare with new validations, robust metadata collection is needed. The Center for Integrated Space Weather Modeling (CISM) started making steps toward this goal by defining a set of operational and scientific quantities to validate each model against (Spence et al., 2004): shock/CME arrival time, speed, B_z , duration at L1 for operational quantities and density, velocity, and Interplanetary Magnetic Field (IMF) for scientific quantities. During this project, just one skill score has been defined that was applied to all quantities, which puts limitations on the metrics. However, the focus of their project remained on assessing the performance of the background solar wind at L1 (see Owens et al., 2008).

To tackle some of the above issues, the Community Coordinated Modeling Center (CCMC) has taken the lead to facilitate the community-wide International Forum for Space Weather Capabilities Assessment starting in 2017 (<https://ccmc.gsfc.nasa.gov/assessment/>). To address the goals of the forum, six physical domains were identified, with multiple working teams within each domain. Two additional teams were established to focus on information architecture and general scientific progress tracking issues common to all the physical domains. Working teams are made up by interested participants from the community and their goals are to take action on defining metrics to assess the current state of space weather modeling capabilities as well as to quantify and track progress over time. Both operational and research needs are taken into account.

This paper focuses on the progress made by the CME Arrival and Impact working team since April 2017 to present. The working team has made considerable efforts in communicating with the research and user community by e-mail, informal discussions, and conference sessions. This includes sessions at the International CCMC (ICCMC)- Living With a Star (LWS) working meeting in April 2017, European Space Weather Week in November 2017, and the CCMC workshop in 2018.

Table 1*Overview of All Currently Active Models Within the CME Arrival and Impact Team With Their Corresponding Developers and Point of Contact*

| Model name | Model developers/points of contact |
|----------------|--------------------------------------|
| COIN-TVD MHD | X. Feng, F. Shen |
| DBM/DBEM | B. Vršnak, M. Dumbović, J. Čalogović |
| EAMv2 | E. Paouris |
| EEGGL+AWSoM | I. Sokolov, W. Manchester, M. Jin |
| EIEvo/EIEvoHI | C. Möstl, T. Amerstorfer |
| Enhanced DBM | P. Hess, J. Zhang |
| EUHFORIA | J. Pomoell, C. Verbeke |
| SARM | Núñez |
| WSA-ENLIL+Cone | D. Odstrčil |

The CME Arrival and Impact working team aims to evaluate how well different models and techniques can predict CME arrival time and impact for a set of predetermined events, based on an agreed-upon time period, with open communication with the community. To reach this goal, the team aims to quantify and establish a set of metrics together with a CME validation event set which will provide a benchmark against which current and future models can be assessed. The working team effort will result in an online database of observations, CME parameter inputs as well as the model inputs/outputs and their corresponding set of metrics. This last effort is in close collaboration with the Information Architecture for Interactive Archives (IAIA) working team and CCMC's web-based validation system Comprehensive Assessment of Models and Events using Library Tools (CAMEL). The work is complementary to the CME Scoreboard (Riley et al., 2018), beginning in 2013 with the goal to collect and display real-time CME predictions and to facilitate the validation of these predictions. The CME Arrival Time Team also collaborates closely with the IMF B_z and 3-D CME Kinematics and Topology working teams. Updates and current status on the CME arrival and impact team can be found at the working team webpage (<https://ccmc.gsfc.nasa.gov/assessment/topics/helio-cme-arrival.php>).

In this paper, we first present current participating models in section 2 and we consider the need for metadata and the team effort regarding the collection of metadata in section 3. In section 4, the proposed event set, based on a combination of two different time periods, is discussed. Metrics are discussed in section 5, where we focus mainly on the contingency table and CME arrival metrics for hits. Last, in section 6, we present a short overview of current on-going community projects related to the metrics and validation of CME arrival time prediction and modeling and in section 7, a brief summary.

2. Models

In this section, we give a brief overview of the models that are currently participating in this community effort, some of which are also participating in the CME Arrival Time Scoreboard. More models can be added in the future, and everyone is welcome to join. Over the past years, many different models have been developed that are able to simulate the propagation of CMEs in the inner heliosphere. These models estimate the arrival time of the CME as well as other important quantities such as the arrival speed. Some models focus on the CME shock for predicting the CME arrival time while others focus on the ejecta. A detailed discussion of the types of models can be found in Zhao and Dryer (2014). Models are categorized in the following: Empirical models, expansion speed models, drag-based models, physics-based models, and time-dependent MHD models, which are also physics based. Below we provide brief descriptions of each model, following the same categories, focusing on the model that participated in the community effort up until now. All of the models and the corresponding model developers and point of contact can be found in Table 1. One last type of CME arrival type prediction our team will consider, not considered in Zhao and Dryer (2014), is any operational forecast in which the the outputs of one or more models are first interpreted by a human before a final forecast is made. Forecasters will adjust model outputs before making their prediction based on general forecaster experience including CME fit confidence, analysis of the available coronal data, effects from compound CME events, and recent performance of the background solar wind model (Riley et al., 2018).

2.1. Empirical Models

2.1.1. Effective Acceleration Model

The Effective Acceleration Model (EAMv2; Paouris & Mavromichalaki, 2017b) is a statistical shock arrival time prediction model that uses an empirical relation between the unprojected speed u_r of the CME and the constant acceleration α of the CME. This model uses the calculated “effective acceleration” using data from a new ICME catalogue, covering the years 1996–2009 (Paouris & Mavromichalaki, 2017a), making the hypothesis that the ambient solar wind interacts with ICMEs with a constant acceleration (or deceleration). The unprojected CME radial speed u_r is determined from the speed in the plane of sky u_0 by using the heliographic coordinates of the associated solar flare. Finally, the speed of the shock and the arrival time at Earth can be determined by using basic physics.

2.1.2. Shock Arrival Model

The Shock Arrival Model (SARM; Núñez et al., 2016) predicts the shock arrival time of CMEs for distances as close as 0.72 AU from the Sun, up to 8.7AU. It uses both CME (radial, earthward or plane-of-the-sky speeds) and flare (peak flux, duration and location) data as inputs. It uses a data set of shocks to calibrate the model's differential equation that is based on aerodynamic drag. The calibration has been performed by optimizing the Mean Absolute Error of the CME arrival time, normalized to 1 AU.

2.2. Drag-Based Models

2.2.1. Drag-Based (Ensemble) Model

The Drag-Based Model (<http://oh.geof.unizg.hr/DBM/dbm.php>) (DBM; Vršnak et al., 2013) describes the propagation of CMEs by assuming that at a certain distance from the Sun, the dynamics that mainly govern the evolution and propagation of the CME in the inner heliosphere are solely dependent on the interaction of the CME with the ambient solar wind. The interaction is then considered by the aerodynamic drag acceleration as a quadratic dependence on the CME relative speed. This allows for the equation of motion to be solved analytically and offers a very fast application to predict arrival time and impact speed of CMEs. In Dumbović et al. (2018), an updated version of DBM, Drag-Based Ensemble Model (DBEM) is presented. It includes ensemble modeling to provide a distribution of possible CME arrival times and speeds. Because the DBM model is computationally inexpensive, DBEM can model a large set of runs using the uncertainty ranges of the CME input parameters. From the outputs, a most likely CME arrival time and speed can then be determined, including prediction uncertainties and a forecast confidence.

2.2.2. Enhanced Drag-Based model

The Enhanced Drag Based model as presented in Hess and Zhang (2015) uses the same assumptions as the DBM/DBEM model as explained in section 2.2.1. However, instead of keeping the drag parameter that determines the aerodynamic drag constant, as is done for DBM/DBEM, it varies this parameter, based on observations, and more specifically CME height-time measurements. This way the model can insert more drag closer to the Sun and decrease the drag parameter gradually as the distance from the Sun increases as measured in the CME height-time plots.

2.2.3. EIEvo/EIEvoHI

The Ellipse Evolution (EIEvo; Möstl et al., 2015) assumes an elliptical CME leading edge with a half-width and aspect ratio. It then uses DBM (see section 2.2.1). From this, it can provide arrival times and arrival speed for any position in the inner heliosphere. The Ellipse Evolution Heliospheric Imager (HI) model (EIEvoHI; Amerstorfer et al., 2018) is an updated version of this model that uses only HI observations (in combination with DBM) to determine all of the input parameters for the elliptical CME model.

2.3. MHD Models

We discuss a total of three MHD-based models: Two of them are currently used in an operational setting while the last one is a model that requires significantly more computational power and is currently only used for scientific research.

2.3.1. WSA-ENLIL+Cone Model

The WSA-ENLIL+Cone model is a background solar wind and CME propagation model that is used by multiple operational space weather agencies worldwide. Multiple versions of the model have been used over the past decade. It consists of two parts: the Wang-Sheeley-Arge (WSA) semiempirical coronal model (Arge & Pizzo, 2000) and the 3-D MHD ENLIL numerical model (Odstrčil et al., 2004). The WSA model uses a photospheric magnetogram as input data and a semiempirical approach to approximate the background solar wind at $21.5 R_{\odot}$, which is the inner boundary for the ENLIL model. At this point in the solar system the background solar wind is supersonic. The ENLIL model uses the results from the WSA coronal model to simulate the background

solar wind and can insert multiple CMEs into the background at $21.3 R_{\odot}$. Different methods can be used to estimate the 3-D CME kinematic and geometric CME input parameters (see section 1). Typically, a CME is inserted in the WSA-ENLIL+Cone model as slices of a homogeneous spherical plasma cloud that has a uniform density, temperature and velocity, with no flux-rope structure. Recently, a spheromak version of ENLIL is being developed to approximate the magnetic field in CMEs.

2.3.2. EUHFORIA

The European Heliospheric FORecasting Information Asset (EUHFORIA; Pomoell & Poedts, 2018) model is a newly developed 3-D MHD model that is currently being transitioned into operations at the the Royal Observatory of Belgium, Belgium. It uses its own version of the WSA model as explained in section 2.3.1 to obtain its inner boundary conditions. The background solar wind parameters as obtained from the WSA model are then used as input at the inner boundary of the MHD model at $21.5 R_{\odot}$. The 3-D time-dependent MHD model then relaxes the background solar wind until a steady state is achieved and on top of the background solar wind simulations, multiple CMEs can be modeled and simulated. Typically, the CME is injected as a dense plasma sphere with a constant radius as explained in Pomoell and Poedts (2018). However, also other similar CME models have been implemented and a discussion can be found in Scolini et al. (2018). The main focus of the model development at this moment is to implement a spheromak CME model that includes a magnetic field structure.

2.3.3. EEGGL+AWSOM

The Alfvén-Wave driven Solar wind Model (AWSOM) is a part of the Space Weather Modeling Framework (SWMF; Jin et al., 2017; Sokolov et al., 2013; van der Holst et al., 2014). AWSOM uses the solar corona and inner heliosphere components of the SWMF that are based on the Block-Adaptive-Tree-Solarwind-Roe-Upwind-Scheme (BATS-R-US) MHD code. The coronal component uses a nonuniform spherical grid extending from the chromosphere to $24 R_{\odot}$. The AWSOM model solves the MHD equations with separate ion and electron temperatures and two equations for the Alfvén wave turbulent energy densities propagating along and counter to the magnetic field lines. Model inputs include a magnetogram and parameters from the Eruptive Event Generator Gibson-Low (EEGGL) flux rope model (Jin et al., 2017). CME eruptions can be modeled low in the solar coronal and propagated to the inner heliosphere.

2.3.4. COIN-TVD MHD

COIN-TVD MHD is a model that simulates the background solar wind using a 3-D COIna INterplanetary Total Varition Dimishing (COIN-TVD) scheme in a sun-centered spherical coordinate system (Feng, 2005; Feng et al., 2003; Shen et al., 2007). The time-dependent 3-D ideal MHD equations include solar rotation (Shen et al., 2007) and a heating source term (Feng et al., 2010; Zhou et al., 2012). The inner boundary of the model starts at $1 R_{\odot}$. First, the model solves for a steady background solar wind. The CME is modeled as a magnetic blob with its center sitting at $r = 5 R_{\odot}$ (see Chané et al., 2006; Shen et al., 2011, 2013).

3. Metadata

Metadata is an important aspect when it comes to the metrics and validation of models. First, it allows users and researchers to know exactly which model, model version, and model settings have been used to obtain the results. Second, it allows for the reproducibility of research results. Reproducibility is emerging as a necessary element that most scientific journals require from their authors and the CME Arrival Time and Impact working team is aiming high to achieve this goal for the community. Including metadata for all validation efforts will also remove the need to store all model outputs as it will allow for runs to be reproduced easily and to store the most relevant outputs only. The recent CME scoreboard study, presented by Riley et al. (2018), made it indeed clear that the future collection of metadata is needed, for example, keeping track of which forecasts are official, and what inputs were used.

As the CME Arrival and Impact working team is collecting a large set of data from 3-D CME input parameters to simulation inputs and outputs, and finally validation results, the need for different sets of metadata is clear. We can distinguish between the following components of metadata: metadata regarding the determination of 3-D CME parameters used as model inputs or to determine model inputs, model description metadata, model input metadata, metadata regarding CME arrival observations, model output metadata, and metadata corresponding to the final validation metrics. In the next section we will discuss each of these metadata components, that together will contain all of the metadata information that is needed to achieve our validation goals. Each metadata component will have a corresponding metadata template file and will be stored following the international standard Space Physics Archive Search and Extract (SPASE) data model (Harvey et al.,

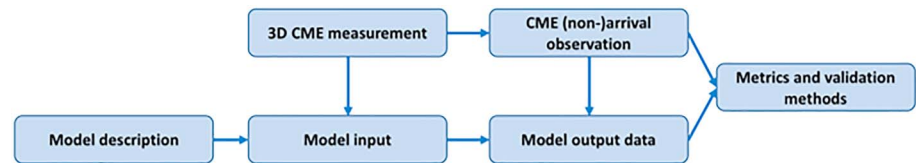


Figure 1. Overview of how each metadata component of the CME Arrival and Impact working team is linked to the others. CME = coronal mass ejection.

2008) when appropriate. In Figure 1, you can find an illustration of how each metadata component is linked to one another.

CCMC is facilitating the storage and linking of metadata with the new CAMEL (<https://ccmc.gsfc.nasa.gov/camel/>) project for web-based validation. Currently, CAMEL is capable of handling time-series metadata but in the near future it will handle all of the metadata types needed for this team. Once model data and metadata from the CME Arrival Time and Impact working team are added to CAMEL, it will be available on the CCMC website for all users. All SPASE metadata template files will be available for all of the metadata components related to this team. In order to define what metadata needs to be collected, the team has also been working closely with operational agencies including the UK Met Office Space Weather Operations Centre (MOSWOC), the Royal Observatory of Belgium (ROB), Solar Influences Data Analysis Center (SIDC), and the National Oceanic and Atmospheric Administration (NOAA) Space Weather Prediction Center (SWPC).

3.1. The 3-D CME Parameter Measurement Metadata

When it comes to the determination of 3-D CME kinematic parameters, such as the CME speed and width, many catalogs and measurement techniques for fitting CMEs are available. These measurement techniques can themselves be considered models which need corresponding metadata. Some publicly available catalogs use automated measurement techniques, such as Computer Aided CME Tracking (CACTUS; Robbrecht et al., 2009), Solar Eruptive Event Detection System (SEEDS; Olmedo et al., 2008), and Coronal Image Processing (CORIMP; Byrne et al., 2012; Morgan et al., 2012), while others rely on CME fitting performed by researchers, such as the CCMC's Space Weather Database Of Notifications, Knowledge, Information (DONKI; ccmc.gsfc.nasa.gov/donki), the SOHO/LASCO catalog (Gopalswamy et al., 2009) and the Heliospheric Cataloguing, Analysis and Techniques Service (HELICATS) HICAT CME catalog (https://www.helcats-fp7.eu/catalogues/wp2_cat.html). While each researcher has their own favorite measurement method, variations in the CME features that are being fit can lead to differences in the obtained 3-D CME parameters. As a result, the working team has come to the conclusion that for all the CME events that are being considered, a new CME fit will be obtained and documented. These new measurements will include images of the model fits so that it is clear to see for every user what exactly has been fit. Vourlidis et al. (2013) and Mays, Thompson, et al. (2015) discusses the importance of separating the CME ejecta from CME-associated brightenings such as streamer deflections and compressive wave fronts from pileups when interpreting the coronagraph images. Currently, we aim to make measurements for only the CME "driver" or "main body" and not the shock front (see Figure 2 of Mays, Thompson, et al., 2015), which is generally the most appropriate input for many of the models. However, the metadata will allow for other fittings to be added in the future, so that model users can choose which part of the CME fit to use, based on the preference of their model. This will reduce the ambiguity regarding the feature that was actually fit to obtain the final 3-D CME measurement parameters. All CME measurements and metadata will be a part of the online CAMEL database. Table 2 summarizes the 3-D CME parameter measurement metadata needed.

The Graduated Cylindrical Shell (GCS) model (Thernisien et al., 2009) has been chosen for the CME measurements. The GCS model uses a forward modeling technique that tries to reproduce a magnetic flux-rope topology and the model shape is similar to a hollow croissant. Model users doing simulation runs for the metrics and validation are allowed to modify the outputs of the GCS model as long as this is done in a consistent way, by, for example, using an algorithm. Some models may need to do this because the GCS model parameters are not corresponding to their model input parameters in a direct way, for example, the GCS model allows for an elliptical cross section of the CME, while some models use spherical shapes. An example of CME fitting using the GCS model can be found in Hess and Zhang (2014).

Table 2
Description of CME and CME Parameter Measurement Metadata

| Name | Required (R) / optional (O) | Description |
|--|-----------------------------|---|
| CME: | | |
| CME onset time | O | The onset time as observed in the associated instrument. |
| Associated instrument(s) | O | e.g., SDO/AIA 193 Å |
| CME start time | R | Timestamp of the first coronagraph observation (associated instrument) of the CME. |
| Associated instrument(s) | R | e.g., SOHO/LASCO C2 |
| Source Location | O | e.g., S17W08 |
| NOAA Active Region Number | O | e.g., 11520 |
| Source Signature Keyword | O | Flare, Double ribbon flare, Post eruption arcade, Filament eruption, Moving/Opening field lines, Brightening, Dimming. |
| Morphology Keyword ^a | O | Flux Rope, Loop, Jet, Other, No Detection, Preceding, Unknown, Shock Candidate. |
| Note Keyword ^a | O | 3-part structure CME, Current sheet, Possible collision with previous CME, Likely deflection of the event within the FOV, CME exhibits dimpled front, Faint event; may affect type assignment, Event fails/disappears before exiting COR2, Front Bright front; may be evidence of pileup, Data gap, Bright emission (likely H-alpha emission), Halo CME, Keyhole-hole shaped CME, Outflowing material at the back of the CME, Event partially overlaps with another CME, Prominence material (filamentary structures), Streamer Blowout following CME, Solar Energetic Particle event, Side-lobe Operations, Surge-like eruption. |
| Note | O | free-form note |
| CME parameters: | | |
| Measurement Technique | R | e.g., GCS, SWPC_CAT, and plane-of-sky |
| Instruments | R | Instruments used for the measurement. |
| Data Level | R | Level of data used for the measurement (e.g., beacon, level 2) |
| Image Type | R | Image processing, e.g., direct, running difference, and base difference. |
| Measurement Feature Code(s) | R | Leading Edge, Trailing Edge, Right Hand Boundary, Left Hand Boundary, Black/White Boundary, Prominence Core, Disconnection Front, Shock Front |
| Image Files | R | Image files showing the CME fit in each measurement frame. |
| Longitude ^b | R | degrees |
| Coordinates | R | e.g., HEEQ |
| Latitude ^b | R | degrees |
| Coordinates | R | e.g., HEEQ |
| Speed ^b | R | CME speed in km/s |
| Height | R | Height corresponding to measured CME speed |
| Time ^b | R | Time (or projected time) |
| Height | R | at a given height (e.g. 21.5 R_{\odot}) |
| Half-width ^b | R | Half of the full CME major angular width (degrees) |
| Minor half-width ^b | O | Half of the full CME minor angular width (degrees) |
| Tilt ^b | O | CME axis/neutral line tilt (degrees counterclockwise from solar equator) |
| Other model specific parameters ^b | O | e.g., GCS parameters κ , α , δ |
| Note | O | free-form note |

^aFollowing the SECCHI CME catalog (Vourlidas et al., 2017). ^bList of multiple measurements for each time stamp may be provided.

Table 3
Brief Description of Model Description Metadata

| Name | Required (R) / optional (O) | Description |
|-----------------------|--------------------------------|--|
| Model name | R | Model Name |
| Model version | R | Model version number |
| Model description | O | Short description |
| Model run description | O | Short description of a typical model run |
| Contacts | R | Model developers/points of contact |
| Publications | R | Short list of most relevant publications |
| Simulation type | O | Type of model as categorized in section 2: empirical, drag-based, MHD-based |
| Region(s) | O | Where is the model valid? e.g., corona and heliosphere |

3.2. Model Description Metadata

As new models are added over time, or current models get updated to a newer version, there is a need to keep track of each model and the changes made to those models. This is achieved by having model description metadata. In Table 3, a list of model description parameters can be found. This metadata contains only a small amount of information but is very crucial for reproducibility as it will be referred to in the model input metadata so that the model and the version number and all relevant general model information is known.

3.3. Model Input Metadata

As mentioned before, reproducibility has become an important part of the scientific community. The model input metadata is needed to be able to reproduce simulation runs. This removes the need to store all model output data, and it is possible to only keep the model output metadata, which requires much less storage space. It is important to also include all relevant model settings, even default ones, that currently may not explicitly be saved. The model settings may change in future model versions and will be tracked this way. This metadata will have to be linked to both the 3-D CME parameter metadata which it is simulating as well as the model output metadata (see Figure 1). In Table 4, a list of input data that is Required (r) and Optional (O) is provided. Note that most of these consist of either a link to other metadata or a list of parameters that are specific to the model. We aim to keep model parameter keywords as general as possible, but some models will have specific model parameters that do not correspond to any parameter in any other model. One example of this can already be found in the differences between WSA-ENLIL and EUHFORIA. While EUHFORIA uses its own coronal model that already calculates all the necessary boundary conditions for the heliospheric model, WSA-ENLIL uses input from WSA, which only provides the speed and magnetic field at the inner boundary of the WSA-ENLIL model. Therefore, the solar wind model parameters of WSA-ENLIL will contain a set of

Table 4
Brief Description of Model Input Metadata

| Name | Required (R) / optional (O) | Description |
|---------------------------------------|--------------------------------|--|
| Model metadata link | R | Link to model description metadata |
| Coronal model metadata link | O | Link to coronal model metadata (if applicable) |
| Solar wind model parameters | R | List with solar wind model parameters (can be empty if not applicable) |
| Model grid parameters | R | List with spatial grid parameters of the model e.g. number of dimensions, grid resolution, inner and outer boundaries |
| Model timing & location parameters | R | List with heliospheric model timing/location parameters (separate from model grid) e.g. relaxation duration, output interval, list of spacecraft/planets for output (can be empty if not applicable) |
| CME input parameters metadata link(s) | R | List with metadata links to all CME inputs (see Table 2) |
| Ensemble runs | O | List with ensemble input parameters (for ensemble models) |

Table 5
Description of CME Arrival Observations Metadata

| Name | Required (R) / optional (O) | Description |
|--------------------------------|-----------------------------|--|
| CME metadata link(s) | R | Link to CME metadata (Table 2) |
| CME parameter metadata link(s) | O | Link to CME parameter measurement metadata (Table 2) |
| CME Arrival: | | |
| Spacecraft | R | Spacecraft/location detecting the arrival |
| Instrument(s) | R | Instrument(s) detecting the arrival |
| Time ^a | R | CME arrival time |
| Signature | R | Shock, sheath, flux-rope, ICME start, ICME end, etc. |
| Criteria | O | Criteria for choosing the time e.g. flux-rope signature. |
| Observer | O | Name of person, or automated algorithm detecting arrival |
| Catalog | O | Associated catalog |
| Type | O | Single, interacting CMEs/compound event, high speed stream |
| <i>Other optional fields</i> | O | e.g., maximum ICME speed, magnetic field, and impact parameter |
| Note | O | free-form note |

^aMultiple time, signature, and criteria triplets may be specified if available.

parameters that for EUHFORIA will be found in the coronal model metadata link. Note that for the coronal model metadata link extra coronal model metadata will have to be created similar to the model input/output metadata as described for the CME Arrival and Impact models.

3.4. CME Arrival Observations Metadata

This metadata contains all information from observations at Earth for a single CME event in the list. Note that for compound events multiple CME parameter metadata will be linked to the same CME arrival metadata, for example, when one CME catches up with a previous CME. One thing that is very important for the working team is to make distinctions between different CME events. We will make distinctions between events such as single CME events, compound events, events with a high speed stream involved, and STEREO-A/B data availability, by the use of flags in the metadata. We will incorporate methods such as those discussed by Zhang et al. (2007) and Kilpua et al. (2017) as guidelines for classifying these types in the metadata. When these are linked to the CME parameter metadata, it also allows users to have access to the CME speed, width, and so forth. This will be important once final metrics are calculated as it will allow the creation of subcategories of the total validation event set to examine if models perform better in certain situations. Model developers can then target certain categories for future model improvement. Table 5 shows proposed metadata fields for describing CME arrival observations.

Table 6
Brief Description of Model Output Metadata

| Name | Required (R)/ optional (O) | Description |
|---|----------------------------|--|
| Model metadata link | R | Link to model description metadata (see Table 3) |
| Model input metadata link | R | Link to model input metadata (see Table 4) |
| CME arrival time | R | List of CME arrival times (+ error bar if applicable) as computed by the model including a link to the metadata of the algorithm used to derive these times |
| CME arrival speed | R | List of CME arrival speeds (+ error bar if applicable) as computed by the model including a link to the metadata of the algorithm used to derive these times |
| CME time-series | R | Link to the time-series produced by the model (can be empty) at all locations |
| Full 2-D/3-D output | O | Link to full 2-D/3-D output of the model |
| Ensemble output | O | Output information specific to ensemble runs |
| CME arrival observational metadata link | C | Link to the CME arrival observational metadata |

Table 7
Description of Metrics and Validation Metadata

| Name | Required (R) / optional (O) | | Description |
|--|-----------------------------|--|---|
| | | | |
| CME metadata link(s) | R | | Link to CME metadata (Table 2) |
| CME arrival observation metadata link(s) | R | | Link to CME metadata (Table 5) |
| Model output metadata link | R | | Link to model output metadata (Table 6) |
| Metric(s) | R | | Definition of each metric (section 5) |
| Threshold(s) | R | | Threshold(s) applicable to each metric |

Over the course of the past years, several CME arrival catalogs have been used such as the Richardson and Cane (2010) ICME catalog (<http://www.srl.caltech.edu/ACE/ASC/DATA/level3/icmetable2.htm>). While most events will be reported by multiple catalogs, each catalog has their own method of deriving the CME arrival time; hence, small differences can be found between catalogs. During the discussions with both researchers and users, the team has decided to take arrival times from the catalogs. If multiple catalogs report the same arrival, then the CME arrival time will be averaged.

3.5. Model Output Metadata

The model output metadata links back to both the model input metadata as well as the CME arrival metadata as seen in Figure 1. The output described by this metadata will be the basis for computing and evaluating model performance and computing metrics. It will contain information about the simulated arrival time of the CME as well as other relevant arrival parameters, such as the simulated peak speed. If the model produces time series, these will also be stored. Also note that it will contain a link to the metadata that contains information about the algorithm used to derive the arrival time of the CME. More specific information about this metadata can be found in Table 6.

3.6. Metrics and Validation Metadata

The metadata for the metrics focuses on describing how we can determine the final computed metrics from all the metadata components above. It will link back to the CME model output metadata, as well as the CME arrival observation metadata. This metadata describes how the contingency table and related scores are created (see section 5.1), and how metrics and skill scores are computed from the model outputs (see section 5.2). This metadata will allow for researchers and users to understand how the metrics are computed and what they represent. Table 7 shows proposed metadata fields for the metrics and validation process.

4. Event Set

4.1. Predetermined Model Inputs for the Event Set

Before discussing the validation time periods as well as the test and training sets, there are a few remarks that need to be made regarding the metadata discussed above. The model input metadata will describe all of the information that is needed to fully reproduce the modeled simulation run, but each model may have a set of free parameters that are related to the CME and solar background wind.

A set of 3-D CME observational parameters will be provided in an online database for each event in the validation study. Each participating model will be required to produce their model results for the validation event set using the provided 3-D CME parameters. Since each model uses a different combination of 3-D CME input parameters, the team will focus on providing the most prominently used parameters, fitting observations to the GCS model. Each modeler can then opt to change these parameters in a consistent way, for example, by the use of an algorithm. After the validation event set has been modeled using the provided (or consistently altered) 3-D CME parameter set, it is possible to also submit model results using their own 3-D CME input parameters, which will be added to the 3-D CME parameters metadata online catalog under a separate keyword. This could be a useful measure of how well a “tuned” or “tweaked” model can do, but this performance will not be compared to other models. However, this is only possible once the model outputs using the provided 3-D CME parameter set have been submitted.

Apart from the 3-D CME related parameters, models typically also have another set of parameters that are related to the computational grid, the background solar wind, and so forth. The working team has agreed that these parameters can be varied throughout the different runs, as long as this is done in a consistent way.

Table 8
List of the 4 Core Events and the Proposed Fitting Parameters, Obtained From Literature

| CME input parameters | Event 1 | Event 2 | Event 3 | Event 4 |
|---|--|--------------------------------|--------------------------------|--------------------------------|
| Event date time (CDAW catalog) | 2010-04-03 10:33 UT | 2013-03-15 07:12 UT | 2014-01-07 18:24 UT | 2015-03-15 01:48 UT |
| Method | GCS | GCS | GCS | GCS |
| Speed (km/s) | (812) | 11.44 | 2565 | 817 |
| HEEQ Longitude (°) | -4 | 9 | 32 | 30 |
| HEEQ Latitude (°) | -26 | -10 | -25 | -15 |
| GCS tilt angle (°) | -1 | -61 | 38 | -22 |
| Height (R_{\odot}) | 13.643 | 8.35 | 11 | |
| GCS ratio | 0.419 | 0.61 | 0.4 | 0.66 |
| GCS half angle (°) | 16.491 | 44 | 40 | 45 |
| Major axis half angle for ellipsoidal model (°) | 34 | 51 | 35 | 52 |
| Minor axis half angle for ellipsoidal model (°) | 26 | 40 | 22 | 43 |
| time at $21.5R_{\odot}$ | (2010-04-03T15:00) | 2013-03-15T10:00 | 2014-01-07T19:44 | 2015-03-15T06:45 |
| Magnetogram | mbqs 2010-04-03T11:54 | mbqs 2013-03-15T07:14 | mbqs 2014-01-07T19:14 | mbqs 2015-03-15T02:04 |
| References | AFFECTS(http://www.affects-fp7.eu/cme-database/cmetadata.php?id=379) | (Beşliu-Ionescu et al., 2014) | (Mays, Thompson, et al., 2015) | (Wang et al., 2016) |
| ACE or Wind arrival time | 2010-04-05T08:26Z ^a | 2013-03-17T05:59Z ^a | 2014-01-09T19:39Z ^a | 2015-03-17T04:45Z ^a |

Note. More details can be found on the working team webpage (<https://ccmc.gsfc.nasa.gov/assessment/topics/CME/events.php>). For Event 1, the parameters between brackets are derived using the SWPC_CAT tool instead of the GCS model fitting. Each magnetogram time stamp corresponds to the first magnetogram after the CME start time. The CME arrival at ACE or Wind is taken from the Richardson and Cane (2010) catalog, except for Event 3 (see Mays, Thompson, et al., 2015).

^a<http://www.affects-fp7.eu/cme-database/cmetadata.php?id=379>.

For example, selecting the magnetogram that has the closest time stamp to the onset of the CME as input, or varying the drag parameter for the background solar wind speed based on another model result or during solar maximum/minimum.

4.2. Validation Time Period and Test Event Set

Most of the current efforts have their own set of events for their validation study (see section 1). One of the key flaws in commonly used validation event sets is that it is not a realistic representation of CME events that a forecaster experiences. The events studied mostly contain all CME hits and events that have clear arrivals from a single CME. In reality there is an overabundance of CMEs that do not arrive at Earth or have flank impacts (see Shen et al., 2014). Therefore, when we assess model performance for predicting CME arrivals, we must also take into account all of the CMEs that never reach Earth. It is for this reason that our team has agreed to not just model a selected set of hit events, but instead to validate a time period consisting of nearly all CME events. This allows for a realistic representation of CME events. Other advantages of this approach are as follows: (1) Modelers can decide to not model a CME and mark it as a nonarrival at Earth based on the CME parameter metadata; (2) modelers can decide to model multiple CMEs together, and the CME parameter metadata will be readily available in the database.

The chosen time period has been split up into two time periods. First, this allows for a time period of both high and low activity, and second, it allows for an extra distinction where we have STEREO-B data available and a period where we do not have this instrument available. For the above reasons, we have decided upon the following two time periods: 1 January 2011 to 31 December 2012 and 1 January 2015 to 31 December 2015. Using the SOHO/LASCO catalog as a starting point, we will down-select the observed CMEs during these time periods using the criteria that the initial speed at $21.5 R_{\odot}$ is greater than 500 km/s and an initial angular width at $21.5 R_{\odot}$ larger than 50° . The SOHO/LASCO catalog reports a total of 475 and 173 CME events for the first and second time period respectively, using the down-selected criteria. For the same time periods, the Richardson and Cane (2010) ICME catalog reports a total of 67 and 30 arrivals at Earth. For the analysis of the hit arrivals to be statistically significant, the working team is aiming for a total of about 100 hit events at Earth, which is in correspondence with the chosen time periods.

In addition to this much larger validation set, which contains both arrivals and nonarrivals at Earth, another much smaller event set is also available. This event set, containing 33 events, focuses on hit events and is part of the NOAA SWPC and CCMC validation effort described in section 6.2. Keep in mind that this event set only contains hit events, and so only the metrics on hit arrivals that are presented in section 5.2 are valid for this validation set (no skill scores).

4.3. Training Event Set

When constructing or improving a model, researchers may use a set of training events to determine how the new model is performing. For this purpose, the working team has provided a small core set of four events, which are generally well studied by the community. Three of those events are a clear hit event, one of which many models predicted as a late arrival. The last event is a false alarm event, where only a weak discontinuity is observed (see section 5.1 for the definition of hit and false alarm). The four core events and proposed fitting parameters, obtained from literature are given in Table 8.

Since many models will use more than those four core training set events, we will of course need to allow for a model's training set to be larger than the core training set. However, since it is very crucial to keep the training set and the validation set separate (especially for those models that use a very large training set), each model will be asked to fully disclose their training set, before starting the validation process. The disclosed training set events will need to be removed from the validation set, if they are overlapping, forming a common validation set. For this reason, we request that the researchers make every effort to select their training set to be outside of the two selected time periods of the validation event set.

5. Metrics and Skill Scores

When it comes to assessing CME arrival predictions, the metrics can be largely divided into two parts: (1) the contingency table and the basic skill scores that can be derived from it, which focuses on all modeled CME events, including those not arriving at Earth, and (2) metrics related to only hits (CME arrival is observed and predicted). While different skill scores have been used by researchers in the past, we are making an effort here to gather all efforts and present a community consensus of the metrics that will be used for model validation

| | | |
|----------------------|------------------|------------------------|
| | Observed Arrival | No Observed Arrival |
| Predicted Arrival | Hit (H) | False Alarm (FA) |
| No Predicted Arrival | Miss (M) | Correct Rejection (CR) |

Figure 2. Overview of the contingency table. Green corresponds to a correct prediction, while red corresponds to a negative prediction.

within the CME Arrival Time and Impact team. We hope that the CME arrival time modeling community will continue using these same metrics in the future, to remove any misconceptions about the metrics. This allows the community to compare models and track improvements over time using a fixed set of metrics. Of course, new metrics and ideas can be implemented over time, while still keeping these standard ones. It is of critical importance when interpreting metrics and skill scores that we understand what they can tell us about model performance and what they cannot. Multiple metrics and skill scores need to be considered in conjunction with each other to get a better overview of model performance. All of the metrics discussed in the sections below are described in detail in Wilks (2011) and Jolliffe and Stephenson (2011).

5.1. Contingency Table

The arrival and/or nonarrival of a CME is a categorical forecast and therefore, a contingency table can be used. The contingency table contains information about how well your model is predicting the (non)arrival of the modeled CMEs. One way to have a consistent representation of a realistic forecasting situation is to model all events over a period of time, as suggested in section 4. In Figure 2, you can find all information contained in a contingency table: hits, misses, false alarms, and correct rejections. In the context of CME arrival, a hit (H) is defined when a CME is both predicted and observed to arrive. A false alarm (FA) is when a CME is predicted to arrive but is not observed to arrive. A miss (M) is when a CME is observed to arrive but it was not predicted to arrive. Finally, a correct rejection (or correct negative; CR) is when a CME is neither predicted nor observed to arrive.

Before we discuss the skill scores that can be calculated from the contingency table, it is important to realize that one has to determine the definition of a hit, for example, within which time frame from an observed CME arrival should a predicted CME arrival be, for it to be counted as a hit in the contingency table. This has been a topic of discussion in our working team, and it became clear that users and scientists have different time intervals that they want to consider. For this reason, the team has recommended that it is useful to analyze performance under different hit definitions. Therefore, we will determine the contingency table for different hit intervals: 3, 6, 12, 18, 24, and 36 hr, where the shorter intervals are more user oriented and the longer time intervals more science oriented.

Now that we have defined the hit definition interval, it is possible to compute the contingency table, based on the model output results. Skill scores can be derived that can tell us more about how well each model

Table 9
Brief Description of Skill Scores Derived From Contingency Tables

| Skill score | Equation | Perfect score | Comments |
|--------------------------------|----------------------|---------------|--|
| Hit rate (POD) | $\frac{H}{H+M}$ | 1 | Fraction of observed arrivals that were predicted. |
| Success Ratio (SR) | $\frac{H}{H+FA}$ | 1 | Fraction of correct predicted arrivals. False Alarm Ratio = 1 – SR |
| Bias Score | $\frac{H+FA}{H+M}$ | 1 | Ratio of predicted arrivals to observed arrivals, < 1 = underforecast; > 1 = overforecast |
| Critical Success Index (CSI) | $\frac{H}{H+M+FA}$ | 1 | Fraction of correct observed arrivals. |
| Accuracy | $\frac{H+CR}{Total}$ | 1 | Fraction of correct forecasts. |
| False Alarm Rate (POFD) | $\frac{FA}{CR+FA}$ | 0 | Fraction of incorrect observed nonarrivals |
| Hanssen & Kuipers discriminant | $HK = POD - POFD$ | 1 | Forecast ability to discriminate between observed event occurrence from nonoccurrence |

Table 10
For Example, Table 3 From Wold et al. (2018) Showing the Hit, Miss, False Alarm, and Correct Rejection Rates for the WSA–ENLIL+Cone Model for the Period March 2010 to December 2016

| | Earth | STEREO-A | STEREO-B | All |
|--------------------|-------|----------|----------|------|
| Hits | 121 | 93 | 59 | 273 |
| False alarms | 180 | 127 | 95 | 402 |
| Misses | 106 | >110 | >75 | >291 |
| Correct rejections | 1293 | 1393 | 1017 | 3703 |

Note. These values are used to compute the POD, SR, Bias, and CSI plotted in the performance diagram of Figure 3.

performs. Our team decided to focus on some basic skill scores well established in the atmospheric sciences. First, we consider a set of four skill scores that are complementary to each other and can be plotted into one figure for clarity, called a performance diagram (Roebber, 2009). These are the probability of detection (POD, also called hit rate), the success ratio (SR), the bias score, and the critical success index (CSI, also known as the threat score). We will also consider two other skill scores: accuracy, which focuses on the fraction of correct forecasts, and the probability of false detection (POFD, also called false alarm rate), which determines the fraction of incorrect observed nonarrivals. From the POD and the POFD, the Hanssen and Kuipers discriminant can be determined as $HK = POD - POFD$ (also known as the true skill statistic or Peirce’s skill score). The HK measures how well the forecast is able to discriminate between two alternative outcomes. All of these skill scores are explained in further detail in Table 9.

To demonstrate a performance diagram for CME arrival forecasts, we have computed these skill scores from the contingency table in Wold et al. (2018), who studied the forecasting capabilities of real-time WSA–ENLIL+Cone simulations of CME arrival time performed at the CCMC from 2010–2016. The corresponding data are summarized in Table 10 and corresponds to Table 3 in Wold et al. (2018). The performance diagram of these scores is shown in Figure 3. The performance diagram has the POD and SR on the axes. The dashed diagonal lines correspond to lines of equal Bias and the full lines correspond to lines of equal CSI. Good performance is when the POD, SR, bias, and CSI approach 1; therefore, a perfect forecast will lie in the upper right corner of the diagram.

Ensemble CME arrival time models produce other outputs in addition to CME arrival time that can also be assessed, such as the likelihood that a CME will arrive. Currently two ensemble models are participating in the team, and we hope more ensemble models will be developed over the years and join the team. The team will begin by using established metrics for probabilistic forecasts that have been applied to CME arrival prediction

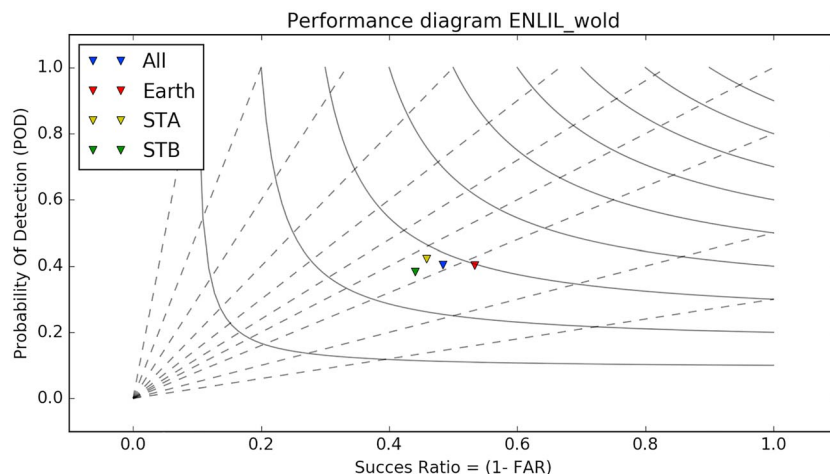


Figure 3. Performance diagram showing the POD, SR, Bias, and CSI skill scores using the contingency table as described in Table 10. The dashed diagonal lines correspond to lines of equal Bias while the solid curves correspond to equal CSIs.

thus far (Dumbović et al., 2018; Mays, Taktakishvili, et al., 2015). Metrics will include the Brier score, Brier skill score, reliability diagram, and rank histogram (Jolliffe & Stephenson, 2011).

5.2. Metrics for Hit Arrivals

5.2.1. Arrival Time

Once we have computed the contingency table and identified the events that are categorized as hits, we can also compute metrics related to the predicted CME arrival time. These are discussed in this section and are based on the metrics that are standard practice in the atmospheric sciences (Jolliffe & Stephenson, 2011). It is important to note that in order to correctly compare the metrics discussed here, each modeled simulation set must be reduced to include the same set of hit events. This “common subset” of hits is an intersection of all hits from each model, as each model is able to have a different set of hit events. This is one of the lessons we learned from validating the CME Arrival Time scoreboard forecasts (Riley et al., 2018). However, the first goal of the CME arrival team is not to compare different models, but to give model developers and users an idea about how each model performs and for which type of events the model performs best. To achieve the secondary team goal of comparing model performance under standardized conditions, error bars must be computed for all of the skill scores and metrics described throughout this section. More information on the error bars on the skill scores can be found at the end of this section.

The time error Δt for one particular forecast, is defined as follows:

$$\Delta t = t^f - t^o, \quad (1)$$

where t^f is the predicted CME arrival time, while t^o represents the observed CME arrival time. We follow the standard practice from atmospheric sciences (Jolliffe & Stephenson, 2011) and the CME Arrival Time scoreboard but this definition has a minus sign difference compared to Riley et al. (2018). With this definition a negative Δt corresponds to when a CME is predicted to arrive earlier than it is observed, while a positive Δt corresponds to a late arrival prediction compared to the observations. Note that there may be different algorithms for determining t^f , which should be specified and differences considered in any validation study.

The first metric is the straight-forward Mean Error (ME) given by

$$ME = Bias = \frac{1}{N} \sum_{i=1}^N \Delta t_i \quad (2)$$

where N is the total number of hit events in the set and i is referring to the event number. The mean error is a way of quantifying the bias, because it explains if the model, on average, is predicting early or late compared to the observed arrivals. A negative bias corresponds to on average early predicted arrivals, while a positive bias corresponds to on average late predicted arrivals.

Next, we consider the most commonly used metric for CME arrival time, the mean absolute error (MAE), defined as

$$MAE = \frac{1}{N} \sum_{i=1}^N |\Delta t_i|. \quad (3)$$

While the ME is a measure of the model's bias, it is not adequate to measure the forecasting skill of a model, since negative errors can compensate positive errors. By taking the absolute value of the errors and in fact measuring the distance between the observed and forecast values, we can compensate for this. The MAE is very similar to the mean square error (MSE) and the root-mean-square error (RMSE), which are discussed later, but does have some differences. In practice, it is used more often and it is also more resistant to outlier errors. Keep in mind that the MAE includes systematic terms that are giving the model a certain bias, as reported by the ME.

Apart from the first-order moments, we can also look at higher order moments. We discuss here only the second-order RMSE, since the MSE does not have the same units as the forecast quantity and the RMSE is the root of the MSE. The RMSE is given by

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\Delta t_i)^2}. \quad (4)$$

Compared to the MAE, the RMSE gives more weight to larger errors, due to the errors being squared. For example, two errors of 1 hr, contribute the same error to MAE as one single error of 2 hr. However, the two 1-hr errors contribute less to RMSE than the one 2-hr error.

One last skill metric that is the measure most commonly used for the spread, is the standard deviation (s.d.) and it is defined as the square root of the variance

$$s.d. = \sqrt{\text{variance}} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\Delta t_i - ME)^2}, \quad (5)$$

where N is the number of events in the set, and ME is the Mean Error. In contrast to the MAE, that measures how far the predictions are (in distance), on average, from the observed values, the s.d. focuses on the spread of the observed values around the Mean Error.

For a fair comparison of model performance it is necessary to compute confidence interval or error bars on the ME, MAE, and any other metrics. The confidence intervals will reflect that the validation event set is only a fraction of the real total population of total CME arrivals in general. This represents how statistically significant the metrics are, which is necessary for intercomparing model performance. The team will also compute confidence interval for the difference between models, the most powerful way of showing statistical significance in forecast verification. The error bars are related to the amount of events in the validation set, and we will follow the procedures as described in Jolliffe and Stephenson (2011) and an example can be found in Wold et al. (2018).

5.2.2. CME Impact

The team has also considered how to quantify model performance on predicting CME “impact” quantities at any location of interest. Many models also predict a velocity of arrival in addition to arrival time. Other simulations can also model the density, magnetic field, and temperature in addition to velocity as a time series at the spacecraft location. One approach proposed by Jian et al. (2017) is to produce scatter plots of the predicted and observed mean and maximum CME impact quantities derived from the predicted and observed time series. The comparison is quantified using the correlation coefficient and the regression slope. The team is still discussing the best way to quantify this comparison and what metrics to use. For example, should the data and prediction be rolling averaged before the comparison? Should the mean, median, maximum or minimum be used? Which spacecraft should be used for the observations?

Models that produce a predicted time series can also be directly compared to the observational time series using well-established continuous time series validation tools, such as those applied to the background solar wind (e.g., Owens et al., 2008). This includes scatter plots, box plots, ME, MAE, RMSE, and correlation coefficient (Jolliffe & Stephenson, 2011; Wilks, 2011). The team will start with these methods and discuss in the future if any more specialized analysis will be needed.

5.3. Solar Wind Background

The space weather research community has made considerable efforts in quantifying how well a model is predicting the background solar wind (Jian et al., 2015, 2016; MacNeice, 2009; Owens et al., 2008). The propagation of CMEs and the topology of the possible corresponding flux-rope can be distorted by the southward magnetic field as well as solar wind stream interaction regions (e.g., Gopalswamy et al., 2009; Jian et al., 2015; Odstrčil & Pizzo, 1999). However, few efforts have been published so far that systematically quantify the influence of the background solar wind on the modeling, propagation, and arrival of CMEs, because this is not a straightforward task. Additionally, both researchers and forecasters have reported that errors in background solar wind modeling impact their CME arrival time predictions. During the working team discussions, it has become apparent that this is something that should be established quantitatively: When modeling the propagation and arrival of a CME, how is your prediction of the solar wind influencing the final results of arrival time and other relevant quantities?

One way to achieve this goal is to simulate a subset of events and vary the background solar wind conditions while keeping the CME input parameters constant. For this a small subset of CME arrival (and nonarrival) events should be determined that include a wide variety of CME arrivals including both direct arrival and flank arrivals. Another way to achieve this goal is to compare model performance for different background solar wind conditions, such as a high speed stream arriving at Earth before/after the CME or steady background

wind. The team will use these methods as a starting point but further discussion is still needed, including on what metrics to use.

6. Relation to Other Community Projects

6.1. CME Arrival Time Scoreboard

The CME Arrival Time Scoreboard (<https://kauai.ccmc.gsfc.nasa.gov/CMEscoreboard/>) provides a central platform for the community to submit their CME arrival time forecasts in real time, quickly view all arrival forecasts at once in real time, and compare forecasting methods when the event has arrived. The philosophy behind the scoreboard is that testing predictive capabilities before event onset is an important element in prototyping forecasting techniques. Since March 2013, there have been 150 registered users and 733 arrival-time predictions using 26 unique prediction methods for 144 CMEs.

Anyone can browse ongoing (“active”) and past CME scoreboard predictions. Registered users may enter their predictions for CME arrival time under “active” CMEs after logging in. At a minimum, each user must enter their predicted CME arrival time and chooses a prediction method from a drop-down menu (<https://swrc.gsfc.nasa.gov/main/cmemodels/>), and the prediction submission time is automatically recorded in the database. Optionally users can enter an error bar for their prediction and a “confidence” level, potentially from an ensemble prediction method or forecaster experience. The confidence probability is an estimate of how confident the forecaster is that the CME will actually arrive at Earth: from 0% to 100%. Another option is to enter the prediction of the resulting geomagnetic storm strength (*Dst* or *Kp*) if the CME arrives at Earth. Super users can enter active CMEs in addition to entering their predictions. Many CMEs are entered by the UK Met Office MOSWOC forecasters as soon as they predict a potential Earth-directed CME. Each CME table also displays the “Average of all Methods” predicted arrival time, which represents a worldwide ensemble forecast. A few days later, if the CME/shock is observed to arrive, this time is entered and added to the CME information. As soon as an observed arrival is entered, the CME becomes inactive and predictions can no longer be entered or changed. For every forecast, the prediction error is automatically computed using equation (1) (“Difference” column), and the lead time is computed as the observed arrival time-forecast submission time. All of the columns of each CME table of the scoreboard are sortable. If the CME is not observed to arrive at Earth, this is added to the CME information and the CME becomes inactive, representing false alarms in the contingency table. There are currently 55 false alarms and 89 hits recorded in the scoreboard, which gives a success ratio of 62%. Because of this specific setup of the CME scoreboard, correct rejections, and misses are hence not archived.

Riley et al. (2018) have carried out the first analysis of the forecasts submitted to the CME Arrival Time scoreboard. This study found mean absolute arrival time errors ranging from 13 to 17 hr, for a subset of models that had the most forecast submissions. The scoreboard “Average of all Methods” ensemble forecast was found to generally perform as well as or slightly better than the other participating models in the CME Arrival Time Scoreboard. Another interesting result was that the most complex physics-based model outperformed other simpler physics-based or empirical models.

Future plans for the CME Scoreboard include: (1) providing an application program interface (API) for data download, (2) accepting automatic XML prediction submissions describing model inputs, (3) capturing CME all-clear predictions, (4) capturing missed CME predictions, and (5) automated validation by linking with CAMEL. In order to accept automated submissions and the link with CAMEL, the scoreboard will adopt the metadata components created by the CME Arrival Time and Impact working team (see section 3).

6.2. NOAA SWPC/CCMC Partnership for Research to Operations Activities

In 2017, NOAA SWPC and CCMC started a new project under an annex to a memorandum of understanding between NASA and NOAA. The purpose of this project is to assess improvements in CME arrival time forecasts at Earth using the Air Force Data Assimilative Photospheric Flux Transport (ADAPT) model driven by data from the Global Oscillation Network Group (GONG; Harvey et al., 1996) ground observatories and feeding output into the coupled WSA-ENLIL model compared to the current operational version of WSA-ENLIL (without ADAPT). The project is performed in close collaboration with model developers Carl Henney, Nick Arge, and Dusan Odstrcil. Currently, SWPC operational forecasts use WSA version 2.2 and ENLIL version 2.6, driven by a single daily updated zero-point uncorrected GONG map. For the purposes of this project WSA version 4.5 and ENLIL version 2.9e will be used.

The community is encouraged to follow the SWPC/CCMC project website (<https://ccmc.gsfc.nasa.gov/annex/>), which always contains up-to-date information on the status of the project. All simulations performed in support of this project are available for download from the project website.

SWPC has selected a set of 33 historical events and 36 CME operational input parameters have been provided (a few events contain 2 CMEs) over the period of three years from 2012 to 2014. They are available from CCMC's public DONKI database via an API in text (https://kauai.ccmc.gsfc.nasa.gov/DONKI/WS/get/CMEAnalysis.txt?startDate=2012-01-01&endDate=2014-12-31&mostAccurateOnly=false&keyword=swpc_annex) or JavaScript Object Notation (JSON; https://kauai.ccmc.gsfc.nasa.gov/DONKI/WS/get/CMEAnalysis?startDate=2012-01-01&endDate=2014-12-31&mostAccurateOnly=false&keyword=swpc_annex) formats.

The SWPC/CCMC project consists of multiple simulation experiments for the entire event set:

- (a) Benchmark: replicating single GONG map driven WSA version 2.2 and ENLIL version 2.6 with ENLIL version 2.9e
- (b) Time-dependent sequence of GONG maps driving WSA version 4.5 and ENLIL version 2.9e.
- (c) For a single event, test simulation of: Time-dependent sequence of GONG maps driving ADAPT, WSA version 4.5, and ENLIL version 2.9e.
- (d) Single GONG map driving ADAPT, WSA version 4.5 and ENLIL version 2.9e.
- (e) Time-dependent sequence of GONG maps driving ADAPT, WSA version 4.5, and ENLIL version 2.9e.

To achieve (a), of the 33 events, a subset of seven events were chosen to fully test that CCMC could replicate the operational ENLIL version 2.6 using ENLIL version 2.9e. For each stage, ENLIL settings will be kept constant, and if it is desired to check the effect of updating settings, this will be performed in substages, so that only one setting is changed at a time. Note that WSA version 4.5 uses different coefficients in the velocity equation compared to version 2.2. After each stage, the performance of the new simulation results will be compared to the benchmark (stage a) and to other previous stages. At the time of this publication we are currently performing simulations for stages (b) and (c). Simulations will be performed using all 12 realizations of ADAPT for stages (c), (d), and (e). For simulations using a time-dependent sequence of input magnetograms, a 6-hr cadence will be used. ENLIL simulations will be performed at medium resolution with a 1-hr 3-D output time step and 1- to 3-min output at locations of interest.

For the purposes of this project, three basic CME arrival time metrics were chosen: ME, MAE, and RMSE. In the future, the data from this study will be available for further comparisons such as comparing the ICME sheath observed mean or max plasma quantities to the simulated quantities. Since all the chosen events are hits, contingency table skill scores will not be considered, however we will keep track of hits that become misses. The observed arrival times used for this validation study were provided by SWPC and are available in the DONKI database.

Methods and lessons learned from the CME Arrival Time and Impact Team will be applied to this SWPC/CCMC project, and vice versa. Additionally, the SWPC/CCMC project's set of operational parameters for 33 events will also be used as a validation test set by the team. This opens up the possibility for other flux transport, coronal, and heliospheric models to test their performance using the same input parameters as the SWPC/CCMC project.

7. Summary

The CME Arrival and Impact team has made significant progress toward the evaluation of different CME arrival time prediction models by using validation time periods, a set of predetermined inputs, and agreed-upon metrics. The working team has so far focused on three main areas: (1) establishing necessary metadata, (2) choosing a meaningful validation time period/event set, and (3) defining metrics and skill scores.

We have shown that the collection of metadata for all components of the modeling process is a crucial part of community validation. The most important reasons are reproducibility, transparency in results, and automation. Different metadata components have been established to describe the (1) initial observations and measurements of the CME, (2) model description, (3) model input, (4) CME arrival observations, (5) model output, and (6) metrics and validation methods.

The metadata describing the CME measurements will include the GCS model fitting information as well as images of the actual fit, for transparency on what CME feature was fitted to obtain the results. The model input metadata includes all the possible input parameters of the model, while the model output metadata describes the output data from those models. The metadata describing the CME arrival observations contain information about the arrival of the CMEs at Earth and other relevant spacecraft. Finally, the metrics metadata will describe the metrics and any assumptions made for the calculation, so that it is clear what procedure was used. All metadata will be linked to each other where needed.

Next, we have discussed the validation time period/test event set and training event set. The validation event set is based on the following time period: 1 January 2011 to 31 December 2012 and 1 January 2015 to 31 December 2015. Nearly all recorded CME events within this time period will be considered as part of the initial validation set. This will allow for a more realistic representation of CMEs hitting and missing Earth. However, we have made a clear distinction between the validation and the training set, as it is crucial that the two do not overlap. Because of this, a common final validation event set will be created, where any model's training events that are also part of the initial validation set will be removed. It is for this reason that everyone is encouraged to select training events that lie outside of the considered time periods.

The team will provide a set of 3-D CME observational parameters using the GCS technique in an online database for each event in the validation study. We will allow each modeler to modify model input data only if it is done in a transparent and consistent way, to avoid the potential for anyone to optimize their model results submitted to the validation study. Optionally, modelers can also use their own input parameters to provide a useful measure of how well a "tuned" model can do, but this performance will not be compared to the other models in the validation study.

We presented the current set of metrics and skill scores as agreed upon by the community. So far, the team has mainly focused on the arrival time of CMEs and their corresponding metrics. The effects and impact of other quantities as well as the influence of the background solar wind still remains an outstanding issue. The metrics on CME arrival can be divided into two main categories: the contingency table with corresponding scores, and the arrival time metrics that focus on hit arrivals.

The contingency table takes into account hits, false alarms, misses, and correct rejections. From these values, different scores can be computed: POD, SR, Bias, CSI, FAR, Accuracy, POFD, and PSS. The first five scores can be plotted on the performance diagram, to quickly assess the performance of a model. For the metrics on arrival time, which focuses only on hit events, the ME, MAE, RMSE, and s.d. have been chosen, because they complement each other well.

Finally, we want to refer the reader to the team website, where all recent updates on team progress can be found (<https://ccmc.gsfc.nasa.gov/assessment/topics/helio-cme-arrival.php>).

Acknowledgments

C. V. and M. L. M. thank Tara Jensen, Barbara Brown, Neel Savani, and Pete Riley for helpful discussions. The SOHO/LASCO CME catalog (https://cdaw.gsfc.nasa.gov/CME_list) is generated and maintained at the CDAW Data Center by NASA and The Catholic University of America in cooperation with the Naval Research Laboratory. SOHO is a project of international cooperation between ESA and NASA. ICME information was found in the Richardson and Cane (2010) ICME catalog: <http://www.srl.caltech.edu/ACE/ASC/DATA/level3/icmetable2.htm>. We gratefully acknowledge contributions from the model developers and participants of the CME Arrival Time Scoreboard (<http://kauai.ccmc.gsfc.nasa.gov/CMEscoreboard>). C. V. is funded by the Research Foundation - Flanders, FWO PhD fellowship 11ZZ216N. L. K. J. is supported by NASA's Living with a Star program and STEREO mission.

References

- Amerstorfer, T., Möstl, C., Hess, P., Temmer, M., Mays, M. L., Reiss, M. A., et al. (2018). Ensemble prediction of a halo coronal mass ejection using heliospheric imagers. *Space Weather*, *16*, 784–801. <https://doi.org/10.1029/2017SW001786>
- Arge, C. N., & Pizzo, V. J. (2000). Improvement in the prediction of solar wind conditions using near-real time solar magnetic field updates. *Journal of Geophysical Research*, *105*, 10,465–10,480. <https://doi.org/10.1029/1999JA000262>
- Bešliu-Ionescu, D., Mariş Muntean, G., Lăcătuş, D. A., Paraschiv, A. R., & Mierla, M. (2014). Detailed analysis of a GE Oeffective ICME triggered by the March 15, 2013 CME. *Revue Roumaine de Géophysique*, *56*, 41–47.
- Brueckner, G. E., Howard, R. A., Koomen, M. J., Korendyke, C. M., Michels, D. J., Moses, J. D., et al. (1995). The Large Angle Spectroscopic Coronagraph (LASCO). *Solar Physics*, *162*, 357–402. <https://doi.org/10.1007/BF00733434>
- Byrne, J. P., Morgan, H., Habbal, S. R., & Gallagher, P. T. (2012). Automatic detection and tracking of coronal mass ejections. II. Multiscale filtering of coronagraph images. *The Astronomical Journal*, *752*, 145. <https://doi.org/10.1088/0004-637X/752/2/145>
- Chané, E., van der Holst, B., Jacobs, C., Poedts, S., & Kimpe, D. (2006). Inverse and normal coronal mass ejections: Evolution up to 1 AU. *Astronomy & Astrophysics*, *447*, 727–733. <https://doi.org/10.1051/0004-6361:20053802>
- Domingo, V., Fleck, B., & Poland, A. I. (1995). The SOHO Mission: An overview. *Solar Physics*, *162*, 1–37. <https://doi.org/10.1007/BF00733425>
- Dumbović, M., Čalogović, J., Vršnak, B., Temmer, M., Mays, M. L., Veronig, A., & Piantchitsch, I. (2018). The Drag-based Ensemble Model (DBEM) for coronal mass ejection propagation. *The Astrophysical Journal*, *854*, 180. <https://doi.org/10.3847/1538-4357/aaa66>
- Feng, X. (2005). A comparative study on 3-D solar wind structure observed by Ulysses and MHD simulation. *Chinese Science Bulletin*, *50*, 672. <https://doi.org/10.1360/982004-293>
- Feng, X., Wu, S. T., Wei, F., & Fan, Q. (2003). A class of TVD type combined numerical scheme for MHD equations with a survey about numerical methods in solar wind simulations. *Space Science Reviews*, *107*, 43–53. <https://doi.org/10.1023/A:1025547016708>
- Feng, X., Yang, L., Xiang, C., Wu, S. T., Zhou, Y., & Zhong, D. (2010). Three-dimensional Solar WIND Modeling from the Sun to Earth by a SIP-CESE MHD Model with a Six-component Grid. *The Astrophysical Journal*, *723*, 300–319. <https://doi.org/10.1088/0004-637X/723/1/300>
- Fry, C. D., Sun, W., Deehr, C. S., Dryer, M., Smith, Z., Akasofu, S.-I., et al. (2001). Improvements to the HAF solar wind model for space weather predictions. *Journal of Geophysical Research*, *106*, 20,985–21,002. <https://doi.org/10.1029/2000JA000220>

- Gopalswamy, N., Lara, A., Manoharan, P. K., & Howard, R. A. (2005). An empirical model to predict the 1-AU arrival of interplanetary shocks. *Advances in Space Research*, 36, 2289–2294. <https://doi.org/10.1016/j.asr.2004.07.014>
- Gopalswamy, N., Yashiro, S., Michalek, G., Stenborg, G., Vourlidas, A., Freeland, S., & Howard, R. (2009). The SOHO/LASCO CME catalog. *Earth Moon and Planets*, 104, 295–313. <https://doi.org/10.1007/s11038-008-9282-7>
- Gosling, J. T. (1993). The solar flare myth. *Journal of Geophysical Research*, 98, 18,937–18,950. <https://doi.org/10.1029/93JA01896>
- Harvey, C. C., Gangloff, M., King, T., Perry, C. H., Roberts, D. A., & Thieman, J. R. (2008). Virtual observatories for space and solar physics research. *Earth Science Informatics*, 1(1), 5–13. <https://doi.org/10.1007/s12145-008-0008-1>
- Harvey, J. W., Hill, F., Hubbard, R. P., Kennedy, J. R., Leibacher, J. W., Pintar, J. A., et al. (1996). The Global Oscillation Network Group (GONG) Project. *Science*, 272, 1284–1286. <https://doi.org/10.1126/science.272.5266.1284>
- Hess, P., & Zhang, J. (2014). Stereoscopic study of the kinematic evolution of a coronal mass ejection and its driven shock from the sun to the earth and the prediction of their arrival times. *The Astrophysical Journal*, 792, 49. <https://doi.org/10.1088/0004-637X/792/1/49>
- Hess, P., & Zhang, J. (2015). Predicting CME ejecta and sheath front arrival at L1 with a data- constrained physical model. *The Astrophysical Journal*, 812, 144. <https://doi.org/10.1088/0004-637X/812/2/144>
- Howard, R. A., Moses, J. D., Vourlidas, A., Newmark, J. S., Socker, D. G., Plunkett, S. P., et al. (2008). Sun Earth Connection Coronal and Heliospheric Investigation (SECCHI). *Space Science Reviews*, 136, 67–115. <https://doi.org/10.1007/s11214-008-9341-4>
- Hudson, H. S., Bougeret, J. L., & Burkepile, J. (2006). Coronal mass ejections: Overview of observations. *Space Science Reviews*, 123, 13–30. <https://doi.org/10.1007/s11214-006-9009-x>
- Huttunen, K. E. J., Schwenn, R., Bothmer, V., & Koskinen, H. E. J. (2005). Properties and geoeffectiveness of magnetic clouds in the rising, maximum and early declining phases of solar cycle 23. *Annales Geophysicae*, 23, 625–641. <https://doi.org/10.5194/angeo-23-625-2005>
- Jian, L. K., MacNeice, P. J., Mays, M. L., Taktakishvili, A., Odstrcil, D., Jackson, B., et al. (2016). Validation for global solar wind prediction using Ulysses comparison: Multiple coronal and heliospheric models installed at the Community Coordinated Modeling Center. *Space Weather*, 14, 592–611. <https://doi.org/10.1002/2016SW001435>
- Jian, L. K., MacNeice, P. J., Taktakishvili, A., Odstrcil, D., Jackson, B., Yu, H. S., et al. (2015). Validation for solar wind prediction at Earth: Comparison of coronal and heliospheric models installed at the CCMC. *Space Weather*, 13, 316–338. <https://doi.org/10.1002/2015SW001174>
- Jian, L. K., Odstrcil, D., Mays, M. L., & Rouillard, A. (2017). Simulation of CMEs in 2007–2016 Using the WSA-ENLIL+Cone Modeling System. In *14th European Space Weather Week*.
- Jin, M., Manchester, W. B., van der Holst, B., Sokolov, I., Tóth, G., Mullinix, R. E., et al. (2017). Data-constrained coronal mass ejections in a global magnetohydrodynamics model. *The Astrophysical Journal*, 834, 173. <https://doi.org/10.3847/1538-4357/834/2/173>
- Jolliffe, I., & Stephenson, D. (Eds.) (2011). *Forecast verification: A practitioner's guide in atmospheric science* (2nd ed.). New Jersey, USA: Wiley.
- Kaiser, M. L., Kucera, T. A., Davila, J. M., St. Cyr, O. C., Guhathakurta, M., & Christian, E. (2008). The STEREO Mission: An introduction. *Space Science Reviews*, 136, 5–16. <https://doi.org/10.1007/s11214-007-9277-0>
- Kay, C., Gopalswamy, N., Reinard, A., & Opher, M. (2017). Predicting the magnetic field of Earth-impacting CMEs. *The Astrophysical Journal*, 835, 117. <https://doi.org/10.3847/1538-4357/835/2/117>
- Kilpua, E. K. J., Balogh, A., von Steiger, R., & Liu, Y. D. (2017). Geoeffective properties of solar transients and stream interaction regions. *Space Science Reviews*, 212, 1271–1314. <https://doi.org/10.1007/s11214-017-0411-3>
- Koskinen, H. E. J., & Huttunen, K. E. J. (2006). Geoeffectivity of coronal mass ejections. *Space Science Reviews*, 124, 169–181. <https://doi.org/10.1007/s11214-006-9103-0>
- MacNeice, P. (2009). Validation of community models: 2. Development of a baseline using the Wang-Sheeley-Argge model. *Space Weather*, 7, S12002. <https://doi.org/10.1029/2009SW000489>
- Mays, M. L., Taktakishvili, A., Pulkkinen, A., MacNeice, P. J., Rastätter, L., Odstrcil, D., et al. (2015). Ensemble Modeling of CMEs Using the WSA-ENLIL+Cone Model. *Solar Physics*, 290, 1775–1814. <https://doi.org/10.1007/s11207-015-0692-1>
- Mays, M. L., Thompson, B. J., Jian, L. K., Colaninno, R. C., Odstrcil, D., Möstl, C., et al. (2015). Propagation of the 7 January 2014 CME and resulting geomagnetic non-event. *The Astrophysical Journal*, 812, 145. <https://doi.org/10.1088/0004-637X/812/2/145>
- McKenna-Lawlor, S. M. P., Dryer, M., Kartalev, M. D., Smith, Z., Fry, C. D., Sun, W., et al. (2006). Near real-time predictions of the arrival at Earth of flare-related shocks during Solar Cycle 23. *Journal of Geophysical Research*, 111, A11103. <https://doi.org/10.1029/2005JA011162>
- Mierla, M., Inhester, B., Antunes, A., Boursier, Y., Byrne, J. P., Colaninno, R., et al. (2010). On the 3-D reconstruction of coronal mass ejections using coronagraph data. *Annales Geophysicae*, 28, 203–215. <https://doi.org/10.5194/angeo-28-203-2010>
- Millward, G., Biesecker, D., Pizzo, V., & Koning, C. A. (2013). An operational software tool for the analysis of coronagraph images: Determining CME parameters for input into the WSA-Enlil heliospheric model. *Space Weather*, 11, 57–68. <https://doi.org/10.1002/swe.20024>
- Möestl, C., Amla, K., Hall, J. R., Liewer, P. C., Jong, E. M. D., Colaninno, R. C., et al. (2014). Connecting speeds, directions and arrival times of 22 coronal mass ejections from the Sun to 1 au. *The Astrophysical Journal*, 787(2), 119.
- Morgan, H., Byrne, J. P., & Habbal, S. R. (2012). Automatically detecting and tracking coronal mass ejections. I. Separation of dynamic and quiescent components in coronagraph images. *The Astrophysical Journal*, 752, 144. <https://doi.org/10.1088/0004-637X/752/2/144>
- Möstl, C., Rollett, T., Frahm, R., Liu, Y., Long, D., Colaninno, R., et al. (2015). Strong coronal channeling and interplanetary evolution of a solar storm up to Earth and Mars. *Nature Communications*, 6, 7135. <https://doi.org/10.1038/ncomms8135>
- Núñez, M., Nieves-Chinchilla, T., & Pulkkinen, A. (2016). Prediction of shock arrival times from CME and flare data. *Space Weather*, 14, 544–562. <https://doi.org/10.1002/2016SW001361>
- Odstrčil, D., & Pizzo, V. J. (1999). Three-dimensional propagation of coronal mass ejections in a structured solar wind flow 2. CME launched adjacent to the streamer belt. *Journal of Geophysical Research*, 104, 493–504. <https://doi.org/10.1029/1998JA900038>
- Odstrčil, D., Riley, P., & Zhao, X. P. (2004). Numerical simulation of the 12 May 1997 interplanetary CME event. *Journal of Geophysical Research*, 109, A02116. <https://doi.org/10.1029/2003JA010135>
- Olmedo, O., Zhang, J., Wechsler, H., Poland, A., & Borne, K. (2008). Automatic detection and tracking of coronal mass ejections in coronagraph time series. *Solar Physics*, 248, 485–499. <https://doi.org/10.1007/s11207-007-9104-5>
- Owens, M. J., Spence, H. E., McGregor, S., Hughes, W. J., Quinn, J. M., Arge, C. N., et al. (2008). Metrics for solar wind prediction models: Comparison of empirical, hybrid, and physics-based schemes with 8 years of L1 observations. *Space Weather*, 6, S08001. <https://doi.org/10.1029/2007SW000380>
- 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, 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, 180. <https://doi.org/10.1007/s11207-017-1212-2>
- Pomell, J., & Poedts, S. (2018). EUHFORIA: European heliospheric forecasting information asset. *Journal of Space Weather and Space Climate*, 8(A35), 14. <https://doi.org/10.1051/swsc/2018020>

- 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, 189–237. <https://doi.org/10.1007/s11207-010-9568-6>
- Riley, P., Mays, M. L., Andries, J., Amerstorfer, T., Delouille, V., Dumbovic, M., et al. (2018). Forecasting the arrival time of coronal mass ejections: Analysis of the CCMC CME Scoreboard. *Space Weather*, 16, 1245–1260. <https://doi.org/10.1029/2018SW001962>
- Robbrecht, E., Berghmans, D., & Van der Linden, R. A. M. (2009). Automated LASCO CME catalog for solar cycle 23: Are CMEs scale invariant?. *The Astrophysical Journal*, 691, 1222–1234. <https://doi.org/10.1088/0004-637X/691/2/1222>
- Roebber, P. J. (2009). Visualizing multiple measures of forecast quality. *Weather and Forecasting*, 24(2), 601–608. <https://doi.org/10.1175/2008WAF2222159.1>
- Schrijver, C. J., Kauristie, K., Aylward, A. D., Denardini, C. M., Gibson, S. E., Glover, A., et al. (2015). Understanding space weather to shield society: A global road map for 2015-2025 commissioned by COSPAR and ILWS. *Advances in Space Research*, 55, 2745–2807. <https://doi.org/10.1016/j.asr.2015.03.023>
- Scolini, C., Verbeke, C., Poedts, S., Chané, E., Pomoell, J., & Zuccarello, F. P. (2018). Effect of the initial shape of coronal mass ejections on 3-D MHD simulations and geoeffectiveness predictions. *Space Weather*, 16, 754–771. <https://doi.org/10.1029/2018SW001806>
- Shen, F., Feng, X., Wu, S. T., & Xiang, C. (2007). Three-dimensional MHD simulation of CMEs in three-dimensional background solar wind with the self-consistent structure on the source surface as input: Numerical simulation of the January 1997 Sun-Earth connection event. *Journal of Geophysical Research*, 112, A06109. <https://doi.org/10.1029/2006JA012164>
- Shen, F., Feng, X. S., Wu, S. T., Xiang, C. Q., & Song, W. B. (2011). Three-dimensional MHD simulation of the evolution of the April 2000 CME event and its induced shocks using a magnetized plasma blob model. *Journal of Geophysical Research*, 116, A04102. <https://doi.org/10.1029/2010JA015809>
- Shen, F., Shen, C., Wang, Y., Feng, X., & Xiang, C. (2013). Could the collision of CMEs in the heliosphere be super-elastic? Validation through three-dimensional simulations. *Geophysical Research Letters*, 40, 1457–1461. <https://doi.org/10.1002/grl.50336>
- Shen, C., Wang, Y., Pan, Z., Miao, B., Ye, P., & Wang, S. (2014). Full-halo coronal mass ejections: Arrival at the Earth. *Journal of Geophysical Research: Space Physics*, 119, 5107–5116. <https://doi.org/10.1002/2014JA020001>
- Shiota, D., & Kataoka, R. (2016). Magnetohydrodynamic simulation of interplanetary propagation of multiple coronal mass ejections with internal magnetic flux rope (SUSANOO-CME). *Space Weather*, 14, 56–75. <https://doi.org/10.1002/2015SW001308>
- Sokolov, I. V., van der Holst, B., Oran, R., Downs, C., Roussev, I. I., Jin, M., et al. (2013). Magnetohydrodynamic waves and coronal heating: Unifying empirical and MHD turbulence models. *The Astrophysical Journal*, 764, 23. <https://doi.org/10.1088/0004-637X/764/1/23>
- Spence, H., Baker, D., Burns, A., Guild, T., Huang, C.-L., Siscoe, G., & Weigel, R. (2004). Center for integrated space weather modeling metrics plan and initial model validation results. *Journal of Atmospheric and Solar-Terrestrial Physics*, 66, 1499–1507. <https://doi.org/10.1016/j.jastp.2004.03.029>
- Taktakishvili, A., Kuznetsova, M., MacNeice, P., Hesse, M., Rastätter, L., Pulkkinen, A., et al. (2009). Validation of the coronal mass ejection predictions at the Earth orbit estimated by ENLIL heliosphere cone model. *Space Weather*, 7, S03004. <https://doi.org/10.1029/2008SW000448>
- Thernisien, A., Vourlidas, A., & Howard, R. A. (2009). Forward modeling of coronal mass ejections using STEREO/SECCHI data. *Solar Physics*, 256, 111–130. <https://doi.org/10.1007/s11207-009-9346-5>
- Thompson, W. T. (2009). 3D triangulation of a Sun-grazing comet. *Icarus*, 200, 351–357. <https://doi.org/10.1016/j.icarus.2008.12.011>
- van der Holst, B., Sokolov, I. V., Meng, X., Jin, M., Manchester, W. B. IV., Tóth, G., & Gombosi, T. I. (2014). Alfvén Wave Solar Model (AWSolM): Coronal heating. *The Astrophysical Journal*, 782, 81. <https://doi.org/10.1088/0004-637X/782/2/81>
- Vourlidas, A., Balmaceda, L. A., Stenborg, G., & Dal Lago, A. (2017). Multi-viewpoint coronal mass ejection catalog based on STEREO COR2 observations. *The Astrophysical Journal*, 838, 141. <https://doi.org/10.3847/1538-4357/aa67f0>
- Vourlidas, A., Lynch, B. J., Howard, R. A., & Li, Y. (2013). How many CMEs have flux ropes? Deciphering the signatures of shocks, flux ropes, and prominences in coronagraph observations of CMEs. *Solar Physics*, 284, 179–201. <https://doi.org/10.1007/s11207-012-0084-8>
- Vršnak, B., Temmer, M., Žic, T., Taktakishvili, A., Dumbović, M., Möstl, C., et al. (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, 21. <https://doi.org/10.1088/0067-0049/213/2/21>
- Vršnak, B., Žic, T., Vrbanc, D., Temmer, M., Rollett, T., Möstl, C., et al. (2013). Propagation of interplanetary coronal mass ejections: The drag-based model. *Solar Physics*, 285, 295–315. <https://doi.org/10.1007/s11207-012-0035-4>
- Wang, Y., Zhang, Q., Liu, J., Shen, C., Shen, F., Yang, Z., et al. (2016). On the propagation of a geoeffective coronal mass ejection during 15-17 March 2015. *Journal of Geophysical Research: Space Physics*, 121, 7423–7434. <https://doi.org/10.1002/2016JA022924>
- Webb, D. F., & Howard, R. A. (1994). The solar cycle variation of coronal mass ejections and the solar wind mass flux. *Journal of Geophysical Research*, 99, 4201–4220. <https://doi.org/10.1029/93JA02742>
- Wilks, D. (2011). *Statistical methods in atmospheric sciences: An introduction*. Massachusetts, USA: Academic Press.
- Wold, A. M., Mays, M. L., Taktakishvili, A., Jian, L. K., Odstrčil, D., & MacNeice, P. (2018). Verification of real-time WSA-ENLIL+Cone simulations of CME arrival-time at the CCMC from 2010 to 2016. *Journal of Space Weather and Space Climate*, 8, A17. <https://doi.org/10.1051/swsc/2018005>
- Xie, H., Ofman, L., & Lawrence, G. (2004). Cone model for halo CMEs: Application to space weather forecasting. *Journal of Geophysical Research*, 109, A03109. <https://doi.org/10.1029/2003JA010226>
- Xue, X. H., Wang, C. B., & Dou, X. K. (2005). An ice-cream cone model for coronal mass ejections. *Journal of Geophysical Research*, 110, A08103. <https://doi.org/10.1029/2004JA010698>
- Zhang, J., Richardson, I. G., Webb, D. F., Gopalswamy, N., Huttunen, E., Kasper, J. C., et al. (2007). Solar and interplanetary sources of major geomagnetic storms (Dst = -100 nT) during 1996-2005. *Journal of Geophysical Research*, 112, A10102. <https://doi.org/10.1029/2007JA012321>
- Zhao, X., & Dryer, M. (2014). Current status of CME/shock arrival time prediction. *Space Weather*, 12, 448–469. <https://doi.org/10.1002/2014SW001060>
- Zhao, X. P., Plunkett, S. P., & Liu, W. (2002). Determination of geometrical and kinematical properties of halo coronal mass ejections using the cone model. *Journal of Geophysical Research*, 107(A8), 1223. <https://doi.org/10.1029/2001JA009143>
- Zhou, Y. F., Feng, X. S., Wu, S. T., Du, D., Shen, F., & Xiang, C. Q. (2012). Using a 3-D spherical plasmoid to interpret the Sun-to-Earth propagation of the 4 November 1997 coronal mass ejection event. *Journal of Geophysical Research*, 117, A01102. <https://doi.org/10.1029/2010JA016380>