

Verification of real-time WSA–ENLIL+Cone simulations of CME arrival-time at the CCMC from 2010 to 2016

Alexandra M. Wold^{1,2,*}, M. Leila Mays³, Aleksandre Taktakishvili^{4,3}, Lan K. Jian^{5,3}, Dusan Odstrcil^{3,6} and Peter MacNeice³

¹ American University, Physics Department, Washington, DC, USA

² University of Colorado Boulder, Aerospace Engineering Sciences, Boulder, CO, USA

³ NASA Goddard Space Flight Center, Greenbelt, MD, USA

⁴ Catholic University of America, Washington, DC, USA

⁵ University of Maryland College Park, College Park, MD, USA

⁶ George Mason University, Fairfax, VA, USA

Received 3 June 2017 / Accepted 15 January 2018

Abstract – The Wang-Sheeley-Arge (WSA)–ENLIL+Cone model is used extensively in space weather operations world-wide to model coronal mass ejection (CME) propagation. As such, it is important to assess its performance. We present validation results of the WSA–ENLIL+Cone model installed at the Community Coordinated Modeling Center (CCMC) and executed in real-time by the CCMC space weather team. CCMC uses the WSA–ENLIL+Cone model to predict CME arrivals at NASA missions throughout the inner heliosphere. In this work we compare model predicted CME arrival-times to in situ interplanetary coronal mass ejection leading edge measurements at *Solar TERrestrial Relations Observatory-Ahead (STEREO-A)*, *Solar TERrestrial Relations Observatory-Behind (STEREO-B)*, and Earth (*Wind* and *ACE*) for simulations completed between March 2010 and December 2016 (over 1,800 CMEs). We report hit, miss, false alarm, and correct rejection statistics for all three locations. For all predicted CME arrivals, the hit rate is 0.5, and the false alarm rate is 0.1. For the 273 events where the CME was predicted to arrive at Earth, *STEREO-A*, or *STEREO-B*, and was actually observed (hit event), the mean absolute arrival-time prediction error was 10.4 ± 0.9 h, with a tendency to early prediction error of -4.0 h. We show the dependence of the arrival-time error on CME input parameters. We also explore the impact of the multi-spacecraft observations used to initialize the model CME inputs by comparing model verification results before and after the *STEREO-B* communication loss (since September 2014) and *STEREO-A* sidelobe operations (August 2014–December 2015). There is an increase of 1.7 h in the CME arrival time error during single, or limited two-viewpoint periods, compared to the three-spacecraft viewpoint period. This trend would apply to a future space weather mission at L5 or L4 as another coronagraph viewpoint to reduce CME arrival time errors compared to a single L1 viewpoint.

Keywords: MHD / modeling / validation / forecasting / coronal mass ejection (CME)

1 Introduction

The Wang-Sheeley-Arge (WSA) coronal model (Arge & Pizzo, 2000; Arge *et al.*, 2004) coupled with the global heliospheric ENLIL solar-wind model (Odstrčil *et al.*, 1996; Odstrčil & Pizzo, 1999a, b; Odstrčil, 2003; Odstrčil *et al.*, 2004) has been used extensively in space weather operations world-wide. Space weather models provide forecast capabilities that greatly enhance satellite and ground-based observations. It is essential for both model users and developers to

understand the limitations and capabilities of these models. In order to measure model performance, generally the model output is compared to a measurable parameter and skill scores are computed. Results from model verification are also helpful as feedback to the model developers, setting benchmarks for the current state of a model, and determining the usefulness and capabilities of a model for operations.

Previous studies have also assessed aspects of WSA–ENLIL+Cone model performance, typically with a smaller sample size of selected events, and in a non-real-time setting. Taktakishvili *et al.* (2009) studied the performance of the ENLIL+Cone model in modeling the propagation of

*Corresponding author: amtwold@gmail.com

coronal mass ejections (CMEs) in the heliosphere by comparing the results of the simulation with *ACE* satellite observations. They evaluated the results of the ENLIL+Cone model for 14 fast CME events and found more earlier arrival predictions (9 out of 14) than late arrival predictions. The errors on the earlier arrivals were, on the average, larger than those of the late arrival predictions. The average absolute error was approximately 6 h for the total set. Millward *et al.* (2013) assessed 25 CME events during the first year of WSA–ENLIL operations at NOAA Space Weather Prediction Center (October 2011–October 2012) and found an average error of 7.5 h. Mays *et al.* (2015a) assesses the WSA–ENLIL+Cone ensemble modeling of CMEs in the real-time Community Coordinated Modeling Center (CCMC) setting. The ensemble modeling method provides a probabilistic forecast of CME arrival time and an estimation of arrival-time uncertainty from the spread and distribution of predictions, as well as forecast confidence in the likelihood of CME arrival. For 17 predicted CME arrivals, the mean absolute arrival-time prediction error was 12.3 h with an early bias of -5.8 h. Vršnak *et al.* (2014) compares a drag-based model (DBM) to the WSA–ENLIL+Cone model, and shows that the average arrival time error for DBM is about 14 h. They also show that DBM performs similarly to ENLIL during low solar activity periods, but ENLIL performs better as solar activity increases.

In interpreting the results of arrival time error assessment for the WSA–ENLIL+Cone model, it is important to examine all of the factors that can contribute to this uncertainty. Mays *et al.* (2015a) found that the reliability of ensemble CME-arrival predictions was heavily dependent on the initial distribution of CME input parameters, especially speed and width. Millward *et al.* (2013) provides an analysis of the impact that CME input parameters have on ENLILs arrival time predictions, as well as the importance of multi-viewpoint coronagraph imagery in determining these parameters. Millward *et al.* (2013) found that when CMEs are measured from a single viewpoint with cone analysis, it is difficult to establish objectively the correct ellipse that should be applied to a given halo CME. Any uncertainties in the projected elliptical face of the cone were seen to lead to large errors in calculations of cone angle and radial velocity. Millward *et al.* (2013) reports that three viewpoint measuring of CMEs improves the accuracy of measurements, and that for outputs of models like ENLIL to be meaningful, the accuracy of key CME parameters is essential.

Beyond the improvement in arrival time prediction expected when using three coronagraph viewpoints instead of one or two, tracking CME propagation beyond coronagraph field of views leads to improved arrival time predictions. Möstl *et al.* (2014) found that predicting CME speeds and arrival times with heliospheric images gives more accurate results than using projected initial speeds from coronagraph measurements on the order of 12 h for the arrival times. By comparing predictions of speed and arrival time for 22 CMEs to the corresponding interplanetary coronal mass ejection (ICME) measurements at in situ observatories, they found the absolute difference between predicted and observed ICME arrival times was 8.1 ± 6.3 h (RMS = 10.9 h), with their empirical corrections improving their performance for the arrival times to 6.1 ± 5.0 h (root mean square error RMSE = 7.9 h). While error in arrival time predictions decreases with this heliospheric

image tracking, the prediction lead time decreases. Colaninno *et al.* (2013) used a variety of non-real-time methods to evaluate CME arrival-time predictions based on a linear fit above a height of 50 solar radii (R_{\odot}) to multi-viewpoint imaging data analysis only, and found an average absolute error of 6 h for seven out of nine CMEs, and 13 h for the full sample of nine CMEs.

Möstl *et al.* (2014) and Millward *et al.* (2013) mention important factors other than CME parameter inputs that affect prediction quality. The interaction of multiple CMEs (Lee *et al.*, 2013), the lack of ejecta magnetic structure, differences between a direct hit or a glancing blow (Möstl *et al.*, 2015; Mays *et al.*, 2015b), and prediction errors from other model limitations should be considered. Reliable characterization of the ambient solar wind flow is also necessary for simulating transients and CME propagation (Lee *et al.*, 2013; Mays *et al.*, 2015a).

In this article we evaluate the performance of the WSA–ENLIL+Cone model installed at the CCMC and executed in real-time by the CCMC space weather team from March 2010–December 2016. The CCMC, located at NASA Goddard Space Flight Center, is an interagency partnership to facilitate community research and accelerate implementation of progress in research into space weather operations. The CCMC space weather team is a CCMC sub-team that provides space weather services to NASA robotic mission operators and science campaigns and prototypes new models, forecasting techniques, and procedures. The CCMC space weather team began performing real-time WSA–ENLIL+Cone simulations in March 2010, marking the start of our verification period. The CCMC also serves the *CME Scoreboard* website (<http://kauai.ccmc.gsfc.nasa.gov/CMEScoreboard>) to the research community who may submit CME arrival-time predictions in real-time for different forecasting methods.

In Section 2 we provide a brief description of the WSA–ENLIL+Cone model simulations and input parameters. We describe the methodology of our verification study in Section 3. We compute the average error (bias) and the average absolute arrival time error in Section 4.1 and examine the dependence of these errors on CME input parameters, including direction, speed, and width. In Section 4.2 we determine significance of multi-spacecraft observations for deriving CME parameters used to initialize the model. This is done by comparing arrival time errors before and after the *Solar Terrestrial Relations Observatory (STEREO:Kaiser et al., 2008) Behind (B)* communication loss in September 2014, and during *STEREO-Ahead (A)* sidelobe operations, from August 2014–December 2015. In Section 4.3, we report CME arrival hit, miss, false alarm, and correct rejection statistics and skill scores. For Earth arrivals, verification of K_p forecasts are presented in Section 5. Finally, in Section 6 we summarize our results.

2 WSA–ENLIL+Cone model simulations

The WSA–ENLIL+Cone model consists of two parts, the WSA coronal model that approximates solar wind outflow at $21.5 R_{\odot}$, beyond the solar wind critical point, and the ENLIL 3D MHD numerical model that provides a time-dependent description of the background solar wind plasma and magnetic

field into which a CME can be inserted at the inner boundary of $21.5 R_{\odot}$. A common method to estimate the 3D CME kinematic and geometric parameters is to assume that the geometrical CME properties are approximated by the Cone model (Zhao *et al.*, 2002; Xie *et al.*, 2004), which assumes isotropic expansion, radial propagation, and constant CME cone angular width. Generally, a CME disturbance is inserted in the WSA–ENLIL model as slices of a homogeneous spherical plasma cloud with uniform velocity, density, and temperature as a time-dependent inner boundary condition with a steady magnetic field. Three dimensional CME parameters were determined using the Stereoscopic CME Analysis Tool (StereoCAT) (Mays *et al.*, 2015a) and the NOAA Space Weather Prediction Center CME Analysis Tool (CAT) (Millward *et al.*, 2013). For most of the simulations in this study, ENLIL model version 2.7 was used with ambient settings “a4b1”, together with WSA version 2.2. A small subset of simulations prior to May 2011 (101 simulations; 34 hits) were performed with an earlier ENLIL version 2.6 using the ambient setting of “a3b2”. Jian *et al.* (2011) describes the model version differences in more detail.

In this study, the results of the simulations are compared to in situ ICME arrivals near Earth, *STEREO-A* and *STEREO-B* from March 2010 through December 2016. This set includes simulations of about 1,800 CMEs. Coronagraph observations from the *Solar and Heliospheric Observatory (SOHO: Domingo et al., 1995)* spacecraft at L1 ahead of Earth, and also the *STEREO-A* and *B* spacecraft trailing ahead and behind Earth's orbit were used. Since these simulations are all made in real-time by space weather forecasters, many slow CMEs (under 500 km s^{-1}) and CMEs out of the ecliptic plane (narrow CMEs at latitudes $>25^{\circ}$) may not be modeled. Additionally, all of the coronagraph derived CME measurements were derived in real-time, often inferred from just a few data points due to real-time data gaps. Finally, there are many different space weather forecasters with varying levels of experience producing the measurements that were used to initialize the simulations.

As described in Emmons *et al.* (2013) and Mays *et al.* (2015a), for Earth-directed CMEs, WSA–ENLIL+Cone model outputs are used to compute an estimate of the geomagnetic K_p index using the Newell *et al.* (2007) coupling function. Three magnetic field clock-angle scenarios of 90° (westward), 135° (southwestward), and 180° (southward) are calculated to compute the K_p . Verification results of the K_p predictions are presented in Section 5.

3 Verification methodology

The quality of model performance is evaluated by comparing the model output to the observed ICME arrival time. Both the predicted and observed arrival times refer to the arrival of the leading edge of the ICME shock or compression wave and not the magnetic cloud start time (if it exists). The CCMC's publicly available *Space Weather Database Of Notifications, Knowledge, Information (DONKI)* (ccmc.gsfc.nasa.gov/donki) is populated by the CCMC space weather team and contains space weather relevant flares, solar energetic particles, CME parameters, WSA–ENLIL+Cone CME simulations, ICME arrivals, high speed streams, modeled

magnetopause crossings, and radiation belt enhancements. All simulations used in this study were conducted in real-time and recorded in DONKI, while some of the observed ICME arrivals at L1, *STEREO-A*, and *STEREO-B* were added post-event. In order to determine observed ICME arrivals that may have been missing or incorrect in DONKI, existing ICME catalogs were compared to those already documented in the DONKI database. These included the Jian *et al.* (2006, 2011) *Wind/ACE* and *STEREO* ICME catalogs (Jian *et al.*, 2013) (www-ssc.igpp.ucla.edu/forms/stereo/stereo_level_3.html, ftp://stereodata.nascom.nasa.gov/pub/ins_data/impact/level3/README.html), Richardson & Cane (2010), ICME catalog (www.srl.caltech.edu/ACE/ASC/DATA/level3/icmetable2.html), International Study of Earth-Affecting Solar Transients catalog (solar.gmu.edu/heliophysics/index.php/The_ISEST_Master_CME_List), and the Nieves-Chinchilla *et al.* (2016) *Wind* ICME catalog (wind.nasa.gov/fullcatalogue.php) with circular flux rope model fitting. Any CME arrivals at Earth (L1 spacecraft, e.g. *Wind*, *ACE*, and *DSCOVR* – Deep Space Climate Observatory), *STEREO-A*, and *STEREO-B* from the catalogues that had not been documented in DONKI were added to the database. In the case of any disagreement of arrival times between catalogues or DONKI, the in situ data from *ACE*, *Wind*, *DSCOVR*, and *STEREO-A* and *B* was analyzed.

Complications arise in determining in situ ICME arrivals for several reasons including (1) weak arrivals, (2) hybrid events including arrivals of stream interaction regions (SIRs) and CMEs, and (3) CME arrivals with uncertain sources. The first complication occurs, as an example, when a CME has a slow speed and only creates a minimal jump in observed solar wind parameters. In the second case of SIR/CME hybrids, it can be difficult to distinguish the CME arrival time from the jump caused by the SIR in the in situ data. The third complication is likely to arise when there are multiple CMEs predicted to impact the same location at around the same time. The process of matching observed ICMEs to their source eruptions can be difficult.

Simulation information was retrieved from DONKI with a web-service application program interface (API). The DONKI database contains all CCMC space weather team WSA–ENLIL+Cone model runs, each linked to the CME measurement used as model input. Each CME is also linked to its ICME arrival observed at *STEREO-A*, *STEREO-B*, or Earth, verified with the ICME catalogues listed above. Arrivals at Mercury, Venus, and Mars are also recorded but are not used in this study. The predicted arrival times listed in DONKI are automatically obtained where the derivative of the simulated time series of the dynamic pressure crosses a threshold at each location. The forecaster may also add glancing blow CME arrival predictions by manually assessing the simulation contour plots for cases when a CME arrival is not automatically detected from the time series. The API returns a file in Javascript object notation (JSON) format with all simulation and linked information. We developed automated python routines to process the JSON by CME, retrieving the “CME analysis” measurements (velocity, direction, width) and the simulation results for each measurement. We calculated hit, miss, false alarm, and correct rejection statistics based on whether a predicted or observed arrival at Earth, *STEREO-A*, and/or *STEREO-B* is recorded (see Sect. 4.3). For the hits,

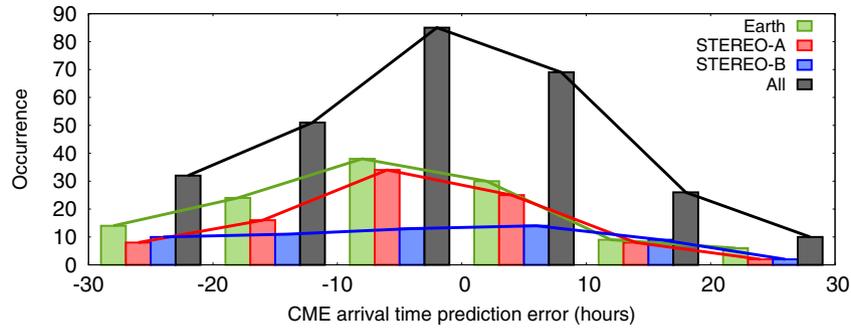


Fig. 1. Distribution of CME arrival time prediction errors at Earth (green), *STEREO-A* (red), *STEREO-B* (blue), and all locations (black). The bins are as follow for each of the locations: [-30, -20], [-20, -10], [-10, 0], [0, 10], [10, 20], [20, 30]. The results for each location are distributed laterally within each bin space for clearer presentation.

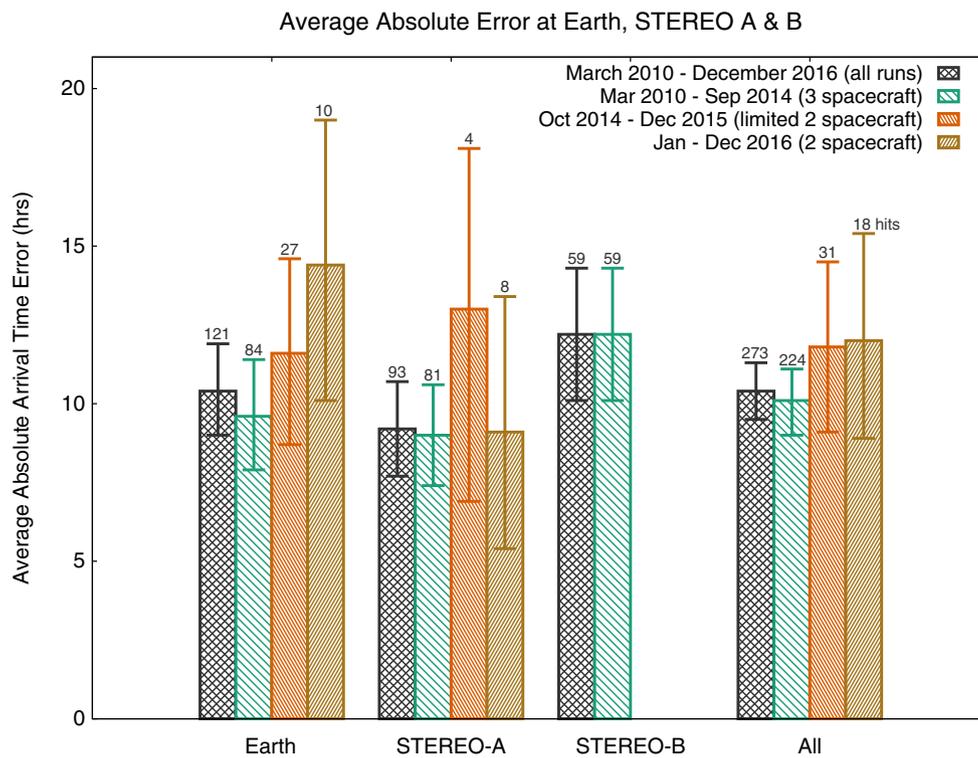


Fig. 2. Average absolute error of CME arrival time predictions at Earth, *STEREO-A*, *STEREO-B*, and all together for four different time periods.

CMEs both predicted and observed to arrive, we calculated the time difference, in hours, of the predicted and observed arrival.

This automated verification is complicated by the fact that multiple simulations are often performed for the same CME, either with updated CME input parameters or in conjunction with other CMEs. Since each CME analysis measurement can be flagged as either “True” or “False” in DONKI to indicate the best measurement, filtering out “False” CME analyses and filtering by most final simulation results for each CME eliminates repeated CMEs from the error calculations. In cases when a CME with unchanging parameters is simulated multiple times (which occurs only when the CME is simulated with other CMEs), we only consider the first simulation of the CME that was performed. This helps eliminate any double counting of the same CME in different simulations that could skew our results.

Another problem can arise for simulations containing multiple CMEs impacting different locations. For example, if two CMEs are modeled together, ENLIL may output a predicted arrival at *STEREO-A* and at Earth. Only a visual examination of the simulation output can clearly show which CME is predicted to impact each location. The WSA-ENLIL+Cone model results stored in DONKI do not contain information that differentiates between which CME is predicted to impact each location, but this is a feature that CCMC plans to add. Since the predicted impacts of two CMEs modeled together are connected directly to the model and not also to their respective CMEs, this introduces some uncertainty to the skill score calculations because some correct rejections and misses may be counted multiple times for the same simulation, and some spurious false alarms will be counted. To

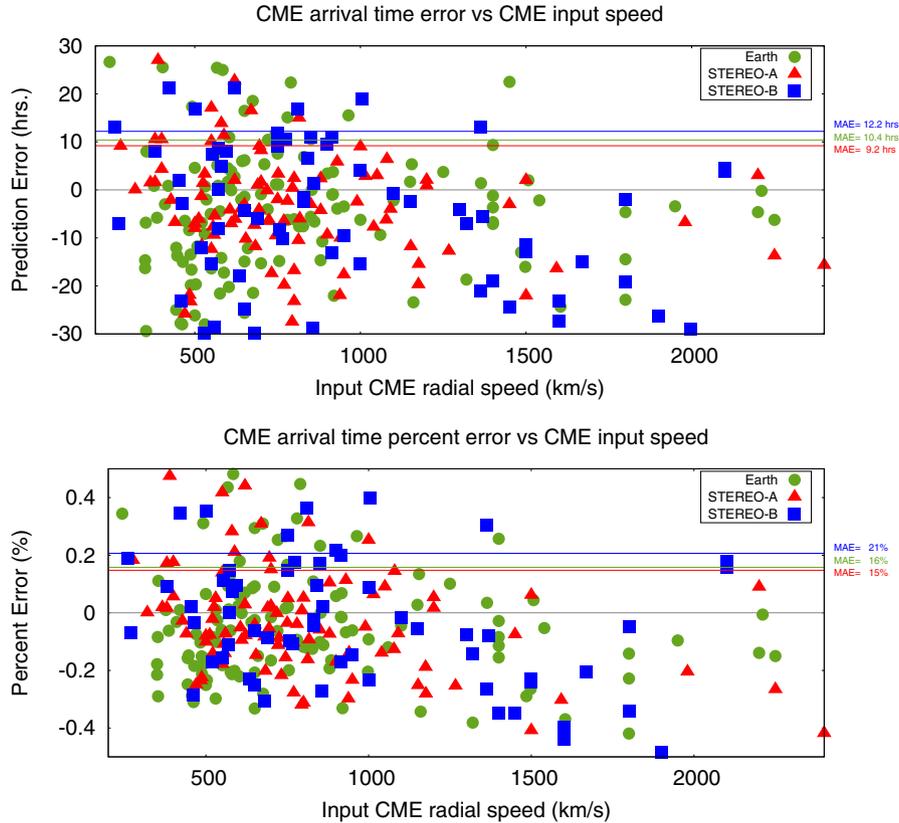


Fig. 3. CME arrival time prediction error versus CME input radial speed (top) and CME arrival time percent error versus CME input radial speed (bottom).

test the impact of this database ambiguity, we excluded multiple simulations including the same CME (40% of our sample) and most of the skill scores remained within the error bars reported (see discussion in Sect. 4.3 and Fig. 6). Note that this uncertainty does not impact our analysis of hits and CME arrival time error, as each CME is directly linked to an observed arrival in the database.

4 CME arrival time verification

4.1 CME arrival time prediction errors

We computed the CME arrival time prediction error $\Delta t_{\text{err}} = t_{\text{predicted}} - t_{\text{observed}}$ for hits at all three locations (*STEREO-A*, Earth, *STEREO-B*). A hit is defined when a CME is predicted and also observed to arrive. In the case of simulations that had arrival time prediction errors greater than 30 h, the simulation was not counted as a hit, but as a miss. Figure 1 shows the histogram distribution of CME arrival time prediction errors in hours, with negative error indicating early prediction and positive error indicating late prediction. Overall at all locations, we found an average arrival time error of -4.0 h showing a tendency for early prediction, and this can be seen in Figure 1. The tendency for early predictions is -4.1 h at Earth, -4.0 h at *STEREO-A*, and -3.9 h at *STEREO-B*. The distribution of errors at *STEREO-B* is flat compared to Earth and *STEREO-A*.

Figure 2 shows the absolute arrival time error at each location and all together for the entire time period (black). We

calculated 95% confidence intervals using a bootstrapping method, resampling 10 000 times with replacement. We found an mean absolute error (MAE) of $10.4^{+1.5}_{-1.4}$ h at Earth, $9.2^{+1.5}_{-1.4}$ h at *STEREO-A*, 12.2 ± 2.1 h at *STEREO-B*, and 10.4 ± 0.9 h at all locations considered together. The slightly increased arrival time error at *STEREO-B* may be due to the model being initialized by the oldest magnetogram information among the three locations (however the error bars at each location overlap). As discussed earlier, prediction errors were computed for hits, defined when the CME arrival time error is less than 30 h. To examine the effect of the threshold used to define the hit, we varied the threshold by decreasing it to 24 and 18 h. As expected, the MAE at all locations decreases to 8.9 ± 0.8 and 7.4 ± 0.7 h, respectively.

In Figure 3 (top), we plot the prediction error in hours against the input CME radial speed in km s^{-1} . Looking at CMEs with input speeds below 1000 km s^{-1} , errors are scattered. However, with CMEs with speeds over 1000 km s^{-1} , most arrival time predictions are early, similar to the results found by Mays *et al.* (2015a). This could be a sign of the modeled CME having too much momentum as defined by a combination of the input speed and half-width (which is related to the modeled CME mass). The overestimation of the modeled CME velocity compared to in situ observed values is also due to the modeled CME having a lower magnetic pressure than is observed in typical magnetic clouds. If we exclude CME input speeds below approximately 700 km s^{-1} the MAE becomes 8.5 h at Earth, 8.4 h at *STEREO-A*, 12.2 h at *STEREO-B*. Figure 3 (bottom) shows the CME arrival time

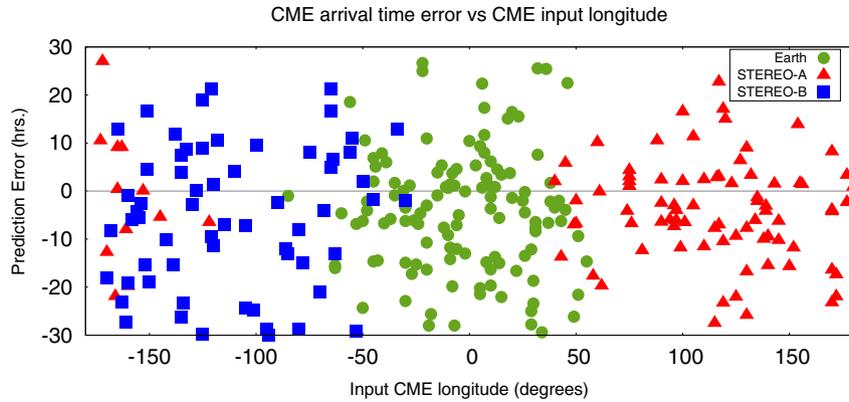


Fig. 4. CME arrival time prediction errors versus CME longitude.

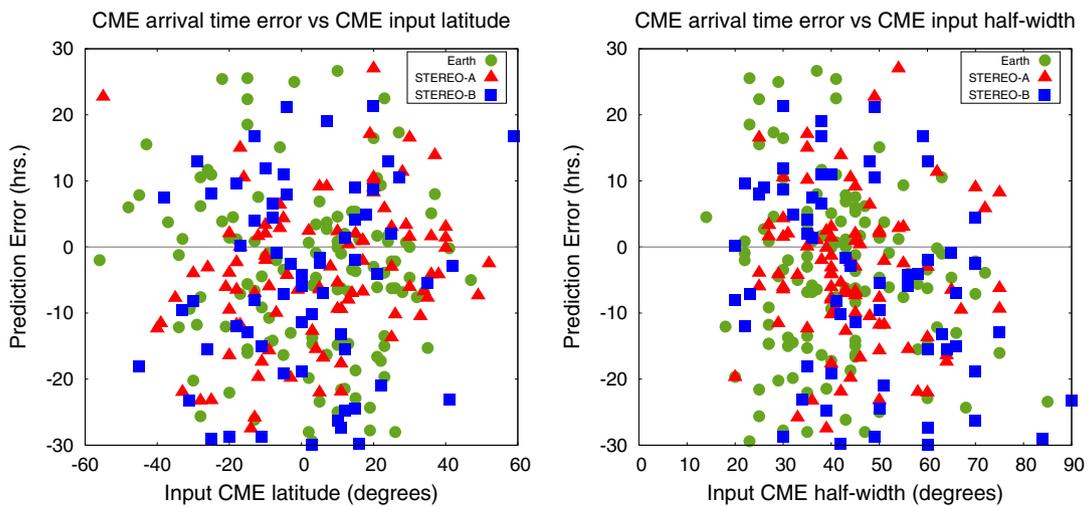


Fig. 5. CME arrival time prediction errors versus input CME latitude (left) and CME half-width (right).

percent error versus the input CME radial velocity. The percent error is calculated as the ratio of the prediction error to the transit time from $21.5 R_{\odot}$ to the detecting spacecraft, and this figure shows similar trends. In investigating any correlation between arrival time error and CME input parameters, no significant correlation was found with latitude, longitude, or width. The distribution of input longitudes in Figure 4 roughly correlates with the detecting spacecraft's location. CME input latitudes in Figure 5 (left) are scattered between $+60^{\circ}$ and -60° , as expected for CMEs that are detected at spacecraft within the ecliptic plane. There is no obvious correlation between error and CME input width in Figure 5 (right).

4.2 Different spacecraft viewpoint time periods

Figure 2 also shows the error at each location for three different time periods, useful for comparing the arrival time errors during periods of two or three observing spacecraft. The full time period of March 2010–December 2016 is shown in black and was discussed in the previous section. The March 2010–September 2014 period had three coronagraph viewpoints observing CMEs, shown in green. *STEREO-A* longitudes ranged from 65° to 168° and *B* -71° to -161° during this time period. After September 2014, communication

with *STEREO-B* was lost, leaving only two coronagraph viewpoints. From October 2014 through December 2015 (orange), just after the *STEREO-B* communication loss, *STEREO-A* underwent sidelobe operations, resulting in very limited *STEREO-A* coverage during this period and effectively only a single spacecraft was available for CME measurement. Normal *STEREO-A* operations resumed in the last time period of January–December 2016 (brown), there are only 18 hits during this period. *STEREO-A* longitudes ranged from -165° to -143° during January–December 2016, which was far from Earth and reduced the effectiveness of dual coronagraph observations.

The average absolute arrival time error at all locations from March 2010–September 2014 with 224 hits (green: three spacecraft) was 10.1 ± 1.0 h. For the 31 hits during the period of *STEREO-A* sidelobe operations and *STEREO-B* communication loss (orange: single-viewpoint in effect), the mean absolute arrival-time prediction error is $11.8^{+2.7}_{-2.6}$ h at all locations. In comparing these two time periods, there is a reduction in skill of $1.7^{+2.9}_{-2.8}$ h. However, due to the small sample size of only 31 hits in the October 2014–December 2015 (orange) *STEREO-A* sidelobe operation period, the 95% error bars overlap. For the 18 hits after the end of *STEREO-A* sidelobe operations, but still without *STEREO-B* communication from January–December 2016 (brown: two spacecraft),

Table 1. Contingency table.

	Observed arrival	No observed arrival
Predicted arrival	Hit (H)	False alarm (FA)
No predicted arrival	Miss (M)	Correct rejection (CR)

Table 2. Brief description of skill scores derived from the contingency table. The false alarm rate is also known as the probability of false detection (POFD) and the hit rate as the probability of detection (POD).

Skill score	Equation	Perfect score	Definition
Success ratio	$H/(H + FA)$	1	Fraction of correct predicted arrivals.
False alarm ratio	$FA/(H + FA)$	0	Fraction of incorrect predicted arrivals.
Accuracy	$(H + CR)/\text{Total}$	1	Fraction of correct forecasts.
Bias score	$(H + FA)/(H + M)$	1	Ratio of predicted arrivals to observed arrivals, <1 = underforecast; >1 = overforecast
Hit rate (POD)	$H/(H + M)$	1	Fraction of observed arrivals that were predicted.
False alarm rate (POFD)	$FA/(CR + FA)$	0	Fraction of incorrect observed non-arrivals
Hanssen and Kuipers discriminant	$HK = \text{POD} - \text{POFD}$	1	Forecast ability to discriminate between observed event occurrence from non-occurrence

Table 3. Hit, miss, false alarm, and correct rejection rates for the WSA–ENLIL+Cone model for the period March 2010–December 2016.

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

the mean absolute arrival-time error is $12.0^{+3.4}_{-3.2}$ h. If we combine the two spacecraft viewpoint periods from October 2014–December 2016 (orange and brown) the error is $11.9^{+2.1}_{-2.0}$ h, with more events (49) reducing the error bar. Compared to the three spacecraft period (green) the difference becomes $1.8^{+2.3}_{-2.2}$ h, still overlapping. The overlapping error bars for these time periods due to the small number of hits in the two-viewpoint time period shows that this result does not have 95% statistical significance. However, if we reduce our bootstrapped confidence interval to 60%, the difference in arrival time errors for three-viewpoint (green) vs. single-viewpoint in effect (orange) time period is 1.7 ± 1.2 h (without overlap). This shows a trend for multi-view coronagraph observations improving CME arrival time forecast accuracy with 60% confidence and that more events are needed to show this with 95% statistical significance.

Similarly, considering only Earth hits, the difference in arrival time errors for the three-viewpoint (green; $\text{MAE} = 9.6 \pm 0.8$ h) vs. single-viewpoint in effect (orange; $\text{MAE} = 11.6 \pm 1.4$ h) time period is 2.0 ± 0.6 h if the confidence interval is reduced to 60% (the 95% confidence interval overlaps). For the 10 hits at Earth from January–December 2016 the $\text{MAE} = 14.4^{+4.5}_{-4.3}$ h. This increased error compared to the sidelobe period could indicate that the *STEREO-A* viewpoint on the farside did not add to greater accuracy in CME measurement. If we combine the two-spacecraft viewpoint periods of October 2014–December 2016 (orange and brown) the error

difference with the three spacecraft period is 2.7 ± 1.9 h (75% confidence interval). When considering only *STEREO-A* arrivals, the difference in arrival time error between the time period with three viewpoints (green) and the time period with an effective single viewpoint (orange) is $4.1^{+4.87}_{-6.7}$ h (with overlap). This large error may suggest that there is a reduction in CME arrival time accuracy when there is only a farside coronagraph (Earth viewpoint) and an extreme ultraviolet instrument (*STEREO-A* viewpoint). However, there are only 4 hits in this reduced time period, so this is not conclusive.

For completeness we have examined all of the CME simulations included in our verification study. This included glancing blow CMEs that are generally more difficult to predict. In the DONKI database predicted glancing blows are manually entered into the system by the forecaster and are flagged as such. We performed the same analysis as Section 4.1 excluding predicted glancing blow arrivals, which reduced the hits at all locations from 273 to 183 and found an average absolute CME arrival time error at all locations of 10.2 ± 1.1 h and $9.5^{+1.7}_{-1.6}$ h at Earth. As expected, the error decreases when glancing blows are excluded, but we only find a slight decrease.

4.3 CME arrival time skill scores

For the purpose of evaluating our forecasting performance, we defined each simulation as a hit, miss, false alarm, or correct

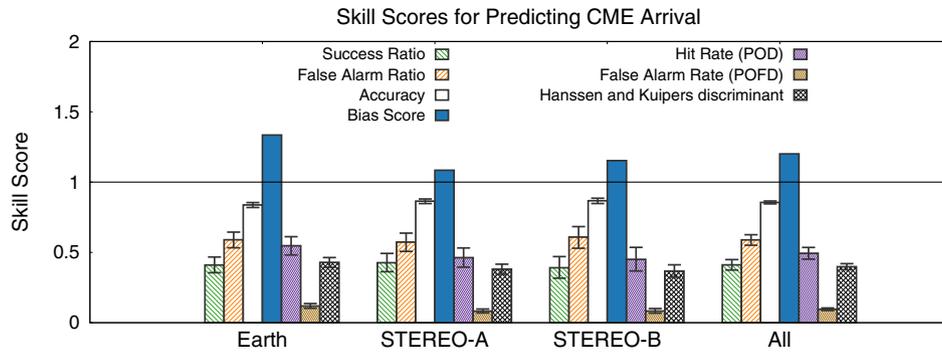


Fig. 6. Success ratio, false alarm ratio, accuracy score, bias score, POD, POFD, and HK (defined in Tab. 2) of total modeled CME events which predict hits at Earth, *STEREO-A*, *STEREO-B*, and all locations combined. Error bars derived from Wilks (2011).

rejection. A hit is an event forecast to occur that did occur. A miss is an event forecast not to occur, but did occur. A false alarm is an event forecast to occur that did not occur. A correct rejection is an event forecast not to occur that did not occur. This is summarized in Table 1. In the case of simulations that had a difference in prediction time and observed arrival time greater than 30 h, the simulation was not counted as a hit, but as a miss. A variety of skill scores were calculated to evaluate model performance as defined in Table 2. The success and false alarm ratios are conditioned on the forecasts (given that an event was forecast, what was the observed outcome?) and represent the fraction of predicted events that were observed, and were not observed, respectively. The accuracy score is the overall fraction of correct forecasts. The bias score indicates if the model is overforecasting or underforecasting by measuring the ratio of the frequency of forecast events to the frequency of observed events, and ranges from -1 to 1 . The probability of detection (POD, or hit rate) and probability of false detection (POFD, or false alarm rate) are conditioned on the observations (given that an event was observed, what is the corresponding forecast?) as the fraction of observed events that were predicted and fraction of incorrect observed non-events, respectively. The Hanssen and Kuipers discriminant ($HK = \text{POD} - \text{POFD}$), or true skill statistic, or Pierce's skill score, ranges from -1 to 1 , with 1 being a perfect score and 0 is no skill. The HK discriminant measures the ability of the forecast to discriminate between two alternative outcomes and does not rely on climatological event frequency (Jolliffe & Stephenson, 2011).

Table 3 shows the number of hits, misses, false alarms, and correct rejections for the real-time simulations run at the CCMC from March 2010 to December 2016. The greater than symbol before the misses at *STEREO-A* and *STEREO-B* indicates that there are a greater number of misses than we could fully confirm at these locations. This is due to the ICME catalogues for *STEREO-A* and *STEREO-B* having not been updated past 2014 at the time of our analysis. They have since been updated through 2016 at the time of this writing.

Figure 6 shows the calculated skill scores discussed in Table 2. With the exception of the bias score, the skill scores do not vary greatly between locations. At all locations over the whole time period, the hit rate, or POD is 0.50 , indicating that half of the observed arrivals were correctly predicted. The false alarm rate, or POFD, is 0.10 and is the fraction of observed non-arrivals that were incorrectly predicted. The Hanssen and Kuipers discriminant ($HK = \text{POD} - \text{POFD}$) of 0.39 represents

how well the prediction is able to separate arrivals from non-arrivals. On the other hand, the success ratio is 0.40 and the false alarm ratio is 0.60 . The success and false alarm ratios represent the fraction of predicted CME arrivals that were observed, and were not observed, respectively. The success ratio is less than the false alarm ratio, and both scores are far from the perfect scores of 1 and 0 , respectively, showing poor skill for these scores which are conditioned on predictions. However, due to spurious false alarms possible from the DONKI database ambiguity discussed in Section 3, when we excluded all simulations of more than one CME (40% of our sample) the success and false alarm ratios become 0.55 and 0.45 respectively, showing some minor improvement. The database ambiguity did not effect any other skill scores to within the error bars shown in Figure 6. The accuracy score at all locations is 0.85 and is closer to the perfect score than the success ratio or hit rate because it is the fraction of correct forecasts overall and the correct rejections are the majority of our forecasts.

The bias score is the ratio of the frequency of forecast events to the frequency of observed events and shows whether the model is underforecasting or over-forecasting. We find a bias score of 1.2 for all locations considered together, revealing a tendency to overforecast CME arrivals. There is a slight bias of 1.1 at *STEREO-A* and *STEREO-B*, but at Earth there is the highest bias for overforecasting with a bias score of 1.3 . This may be due to CCMC space weather team assessing glancing blows in simulation results that were not automatically detected as CME arrivals more often for Earth-directed events, as they have the potential for geoeffectiveness. There is a human bias for over-predicting CME arrivals at Earth by modeling more events. Millward *et al.* (2013) notes that forecaster judgment plays a significant role in event selection and that operational model performance assessments will include the more-difficult-to-forecast glancing blow events compared to non-real-time assessments.

5 K_p prediction verification

The geomagnetic three-hour planetary K_p index (K_p) is derived from data from ground-based magnetometers, categorizes geomagnetic activity on a scale from 0 to 9 , with 0 being the lowest amount of geomagnetic activity and 9 being the highest (Bartels *et al.*, 1939; Rostoker, 1972; Menvielle &

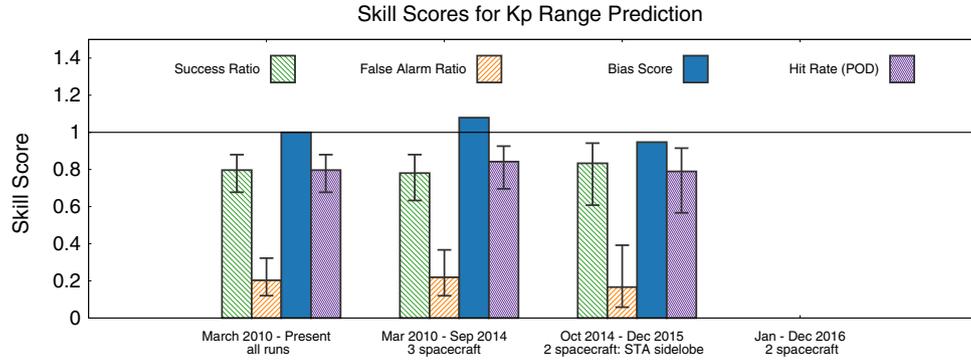


Fig. 7. Success ratio, false alarm ratio, bias score, and hit rate skill scores based on whether the observed K_{pmax} falls within the predicted K_p range, grouped by forecast time period.

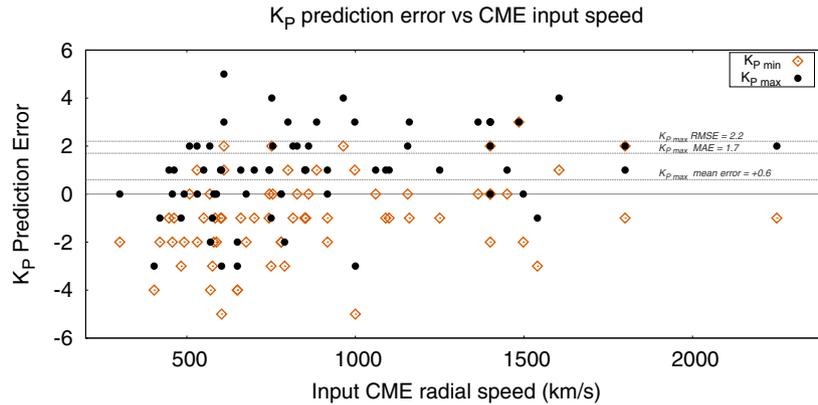


Fig. 8. K_{pmax} and K_{pmin} prediction error plotted against the CME input speed, at times showing large errors given the range of the index from 0 to 9. There is a general overprediction of K_{pmax} for CME input speeds above $\approx 1000 \text{ km s}^{-1}$, whereas the K_{pmin} forecast tends underpredict below $\approx 1000 \text{ km s}^{-1}$.

Berthelier, 1991). The time series that the WSA–ENLIL+Cone model outputs at Earth is used as input to a formula derived from the Newell *et al.* (2007) coupling function to provide a K_p forecast time series for three magnetic field clock-angle scenarios of 90° (westward), 135° (southwestward), and 180° (southward). Because ENLIL-modeled CMEs do not contain an internal magnetic field and the magnetic field amplification is caused mostly by plasma compression, only the magnetic-field magnitude is used and the three magnetic field clock-angle scenarios are assumed. This provides a simple estimate of three possible maximum values of each time series that the K_p index might reach following arrival of the predicted CME shock/sheath. For the forecast, the K_p estimates are rounded to the nearest whole number (Mays *et al.*, 2015a). With all three clock-angle forecasts, we determined a range from the minimum to the maximum predicted K_p and compared to the maximum observed K_p for the three days following the CME arrival. If the maximum observed K_p within three days after the observed CME arrival fell within the predicted K_p range, we counted a hit. If the max observed K_p was above the forecast range, we counted a miss. If the max observed K_p was below the forecast range, we counted a false alarm. For example, if the predicted K_{pmax} for the three clock angle scenarios is 5, 7, 7, and the observed K_p is 6, this is counted as a hit because 6 falls within the predicted K_{pmax} range of 5–7. The resulting

contingency table for K_p range is: 47 hits, 12 false alarms, and 12 misses. There are no correct rejections in this analysis, so we could not calculate an accuracy score. Thus, Figure 7 shows all skill scores from Table 2 except for accuracy, for the different time periods defined in Section 4.1. There were only two K_p predictions for the January–December 2016 time period, so skill scores are not shown. For the entire time period, we found a success ratio of 0.80 and a false alarm ratio of 0.20, meaning that a high fraction of predicted K_p ranges contained observed K_p value and a low fraction of predicted K_p ranges did not contain the observed value. We found a bias score of 1.0, indicating that we are neither under- or over-forecasting the K_p range. The hit rate (POD) of 0.80 shows a high fraction of the observed K_p values were within the K_p range.

The K_p prediction error can be calculated by first taking the maximum of the three K_{pmax} predictions as a single prediction (this is generally 180°). For example, for the earlier prediction of $K_{pmax} = 5, 7, 7$, the single maximum K_{pmax} prediction is 7. For all K_p predictions, the K_{pmax} mean error is 0.6 ± 0.4 , the mean average error is 1.7 ± 0.3 , and the RMSE is 2.2 ± 0.3 . The error bars are computed using a similar bootstrapping method as used for the arrival time error. In Figure 8, K_{pmax} and K_{pmin} prediction errors are plotted against the CME input speed, at times showing large errors given the range of the index from 0 to 9. We find that there is a general

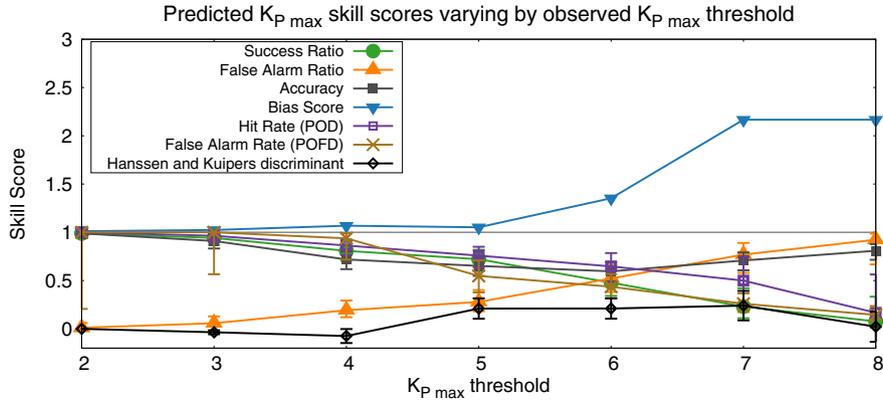


Fig. 9. Success ratio, false alarm ratio, accuracy, bias score, and hit rate skill scores for $K_{p\max}$ forecasts as a function of different thresholds.

overprediction of $K_{p\max}$ for CME input speeds above $\approx 1000 \text{ km s}^{-1}$, similar to the results found by Mays *et al.* (2015a). This overprediction could be due to an overestimation of the CME dynamic pressure at Earth for faster CMEs, as the inserted CMEs have a lower magnetic pressure than is observed and from the approximation of the CME as a cloud with homogeneous density. We also see that the $K_{p\min}$ forecast underpredicts K_p for CME speeds under $\approx 1000 \text{ km s}^{-1}$, likely when the observed K_p is caused by magnetic cloud passage, and not the shock/sheath.

Instead of using the forecast K_p range, a threshold was used to compute the $K_{p\max}$ prediction skill scores in Figure 9. For example, for a threshold of $K_p=6$, a hit occurs when the forecast $K_{p\max}$ and the observed $K_{p\max}$ are at or above 6. A correct rejection is counted when the forecast $K_{p\max}$ and the observed $K_{p\max}$ are both below 6. A false alarm is counted when the forecast $K_{p\max}$ is above or at 6 and the observed $K_{p\max}$ is below 6. A miss is counted when the forecast $K_{p\max}$ is below 6 and the observed $K_{p\max}$ is at or above 6. We varied the threshold between 2 and 8 in Figure 9 and computed the resulting skill scores. The figure shows the success ratio, false alarm ratio, accuracy, bias score, hit rate, false alarm rate, and HK as a function of threshold. The skill scores generally decrease in performance as the $K_{p\max}$ threshold increases, except the false alarm ratio, and HK. With events categorized as hits at low thresholds shifting to correct rejections at higher thresholds, the total number of hits is decreasing. The success ratio, accuracy score, bias score, hit rate, and false alarm rate are all heavily dependent on the number of hits. Only the accuracy score remains relatively unchanged, since it depends on the total number of events (unchanging). This figure shows that the $K_{p\max}$ prediction performs best (where most skill scores perform the best) for an observed $K_{p\max}$ of 5 and above, after which the $K_{p\max}$ is overforecast (bias score) and the false alarm ratio increases.

6 Summary and discussion

In this article we evaluate the performance of the WSA–ENLIL+Cone model installed at the CCMC and executed in real-time by the CCMC space weather team from March 2010 to December 2016. The simulations included over 1,800 CMEs and 273 of these were categorized as hits—CMEs

that were both observed and predicted to arrive. We computed the CME arrival time prediction error $\Delta t_{\text{err}} = t_{\text{predicted}} - t_{\text{observed}}$ for hits at three locations: Earth, *STEREO-A* and *B*. The average absolute error of the CME arrival time was found to be $10.4^{+1.5}_{-1.4}$ h at Earth, $9.2^{+1.5}_{-1.4}$ h *STEREO-A*, 12.2 ± 2.1 h at *STEREO-B*, and 10.4 ± 0.9 h at all locations considered together. These errors are comparable to arrival time error results from other studies using WSA–ENLIL or other models (see Sect. 1). At all locations over the whole time period, the hit rate is 0.50, indicating that half of the observed arrivals were correctly predicted. The false alarm rate is 0.10 and represents the fraction of observed non-arrivals that were incorrectly predicted.

Overall, at all locations, we found an average arrival time error of -4.0 h showing a tendency for early predictions. Sources of CME arrival time error include input CME parameters, ambient solar wind prediction accuracy, previous CMEs, input magnetogram limitations/uncertainties, model ambient parameters, and model CME parameters. For example, Temmer *et al.* (2017) found that interplanetary space takes about 2–5 days to recover to normal background solar wind speed conditions. We are exploring methods to quantify the effect of the ambient solar wind prediction accuracy on CME arrival time, but this is beyond the scope of this article. The tendency for early predictions can also arise from CME input speeds, as measured from coronagraph data, that are too high from measurement of the shock instead of the main CME driver (Mays *et al.*, 2015b). A parametric study by Mays *et al.* (2015a) found that larger CME half widths also have a small tendency to produce early arrivals. They also report that reducing the ENLIL model parameter of CME cloud density ratio [dclld] may help to predict later arrivals. Overall, the parametric case study shows that after the CME input speed, the cavity ratio [radcav] (radial CME cavity width/CME width) and density ratio assumed in ENLIL have the greatest effects on the predicted CME arrival time.

For Earth-directed CMEs using the WSA–ENLIL+Cone model outputs to compute an estimate of the geomagnetic K_p index we found a mean $K_{p\max}$ error of 0.6 ± 0.4 , a mean average error of 1.7 ± 0.3 , and RMSE of 2.2 ± 0.3 , and the $K_{p\max}$ is generally overpredicted for CME input speeds above $\approx 1000 \text{ km s}^{-1}$. The $K_{p\max}$ overprediction could be due to an overestimation of the CME dynamic pressure at Earth for faster CMEs, as the inserted CMEs have no magnetic pressure other than from the ambient field.

Verification of the single-spacecraft (in effect) period October 2014–December 2015 (without *STEREO-B* and with reduced *STEREO-A* coverage) shows an increase in CME arrival time error of $1.7^{+2.9}_{-2.8}$ h. Because of overlapping error bars for these time periods due to the small number of hits in the single-viewpoint (in effect) time period this result does not have 95% statistical significance, but instead 60% (1.7 ± 1.2). Nevertheless we show a trend for multi-view coronagraph observations improving CME arrival time forecast accuracy, and more events are needed to show this with 95% statistical significance. For example, a future space weather mission at L5 or L4 as a second coronagraph viewpoint would reduce CME arrival time errors compared to a single L1 viewpoint. For example, Akioka *et al.* (2005), Simunac *et al.* (2009), Gopalswamy *et al.* (2011), Strugarek *et al.* (2015), Vourlidas (2015), Lavraud *et al.* (2016), and Weinzierl *et al.* (2016) all identify the potential benefit of an L5 mission. Lavraud *et al.* (2016) proposes an L5 mission that would measure the coronal magnetic field using a polarization technique, in addition to white-light imaging. Weinzierl *et al.* (2016) shows how an L5 mission equipped with a magnetic imager could improve predictions of solar wind, which is particularly important for the WSA–ENLIL+Cone model.

We also discussed various factors that should be taken into consideration when interpreting these verification results. This includes uncertainties arising from the determination of CME input parameters from real-time coronagraph data, CME measurements from a variety of forecasters each with their own biases, the identification of ICME arrivals in situ, glancing blow ICME arrivals, and multiple ICME arrivals. In future work, quality factors will be introduced for observed arrivals and identifying candidate CMEs. We will further quantify model performance by evaluating how well observed in situ solar wind measurements compare to modeled values. We will also plan to assess CME arrivals at other locations, such as Mercury (Winslow, 2015) and Mars.

Acknowledgements. This work was completed while A.M.W. was an undergraduate student at American University, now a graduate student at University of Colorado Boulder. A.M.W. and M.L.M. thank Christine Verbeke and the CME Arrival Time and Impact Working Team (ccmc.gsfc.nasa.gov/assessment/topics/helio-cme-arrival.php) for useful discussions. M. L.M. acknowledges the support of NASA grant NNX15AB80G. L.K.J. acknowledges the support of NASA's *STEREO* project, NSF grants AGS 1321493 and 1259549. D. Odstreil acknowledges the support of NASA LWS-SC NNX13AI96G. The WSA model was developed by N. Arge (now at NASA/GSFC), and the ENLIL model was developed by D. Odstreil (now at GMU). Estimated real-time planetary K_p indices are from NOAA as archived on CCMC's iSWA system (ccmc.gsfc.nasa.gov/iswa). The editor thanks Ward Manchester and an anonymous referee for their assistance in evaluating this paper.

References

Akioka M, Nagatsuma T, Miyake W, Ohtaka K, Marubashi K. 2005. The L5 mission for space weather forecasting. *Adv Space Res* **35**: 65–69. Mars International Reference Atmosphere, Living With a Star and Fundamental Physics. DOI:10.1016/j.asr.2004.09.014.

Arge CN, Pizzo VJ. 2000. Improvement in the prediction of solar wind conditions using near-real time solar magnetic field updates. *J Geophys Res* **105**: 10465–10480. DOI:10.1029/1999JA000262.

Arge CN, Luhmann JG, Odstrčil D, Schrijver CJ, Li Y. 2004. Stream structure and coronal sources of the solar wind during the May 12th, 1997 CME. *J Atmos Sol-Terr Phys* **66**: 1295–1309. DOI:10.1016/j.jastp.2004.03.018.

Bartels J, Heck NH, Johnston HF. 1939. The three-hour-range index measuring geomagnetic activity. *Terr Magn Atmos Electr (J Geophys Res)* **44**: 411. DOI:10.1029/TE044i004p00411.

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

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

Emmons D, Acebal A, Pulkkinen A, Taktakishvili A, MacNeice P, Odstrčil D. 2013. Ensemble forecasting of coronal mass ejections using the WSA–ENLIL with CONED Model. *Space Weather* **11**: 95–106. DOI:10.1002/swe.20019.

Gopalswamy N, Davila J, Cyr OS, Sittler E, Auchère F, *et al.* 2011. Earth-Affecting Solar Causes Observatory (EASCO): a potential international living with a star Mission from Sun–Earth L5. *J Atmos Sol-Terr Phys* **73**: 658–663. DOI:10.1016/j.jastp.2011.01.013.

Jian L, Russell CT, Luhmann JG, Skoug RM. 2006. Properties of interplanetary coronal mass ejections at one AU during 1995–2004. *Sol Phys* **239**: 393–436. DOI:10.1007/s11207-006-0133-2.

Jian LK, Russell CT, Luhmann JG, Galvin AB, Simunac KDC. 2013. Solar wind observations at STEREO: 2007–2011. *Sol Wind* **13**: 1539: 191–194. DOI:10.1063/1.4811020.

Jian LK, Russell CT, Luhmann JG, MacNeice PJ, Odstrčil D, Riley P, Linker JA, Skoug RM, Steinberg JT. 2011. Comparison of observations at ACE and Ulysses with Enlil model results: stream interaction regions during Carrington rotations 2016–2018. *Sol Phys* **273**: 179–203. DOI:10.1007/s11207-011-9858-7.

Jolliffe I, Stephenson D, 2011. Forecast verification: a practitioner's guide in atmospheric science, 2nd edn, Wiley, New Jersey, USA.

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

Lavraud B, Liu Y, Segura K, He J, Qin G, *et al.* 2016. A small mission concept to the Sun–Earth Lagrangian L5 point for innovative solar, heliospheric and space weather science. *J Atmos Sol-Terr Phys* **146** (Supplement C): 171–185. DOI:10.1016/j.jastp.2016.06.004.

Lee CO, Arge CN, Odstrčil D, Millward G, Pizzo V, Quinn JM, Henney CJ. 2013. Ensemble modeling of CME propagation. *Sol Phys* **285**: 349–368. DOI:10.1007/s11207-012-9980-1.

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

Mays ML, Thompson BJ, Jian LK, Colaninno RC, Odstrčil D, *et al.* 2015b. Propagation of the 7 January 2014 CME and resulting geomagnetic non-event. *Astrophys J* **812**: 145. DOI:10.1088/0004-637X/812/2/145.

Menvielle M, Berthelier A. 1991. The K-derived planetary indices – description and availability. *Rev Geophys* **29**: 415–432. DOI:10.1029/91RG00994.

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

Möstl C, Amla K, Hall JR, Liewer PC, De Jong EM, *et al.* 2014. Connecting speeds, directions and arrival times of 22 coronal mass

- ejections from the Sun to 1 AU. *Astrophys J* **787**: 119. DOI:[10.1088/0004-637X/787/2/119](https://doi.org/10.1088/0004-637X/787/2/119).
- Möstl C, Rollett T, Frahm R, Liu Y, Long D, et al. 2015. Strong coronal channeling and interplanetary evolution of a solar storm up to Earth and Mars. *Nat Commun* **6**: 7135. DOI:[10.1038/ncomms8135](https://doi.org/10.1038/ncomms8135).
- Newell PT, Sotirelis T, Liou K, Meng C-I, Rich FJ. 2007. A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables. *J Geophys Res* **112**: A01206. DOI:[10.1029/2006JA012015](https://doi.org/10.1029/2006JA012015).
- Nieves-Chinchilla T, Linton MG, Hidalgo MA, Vourlidas A, Savani NP, Szabo A, Farrugia C, Yu W. 2016. A circular-cylindrical flux-rope analytical model for magnetic clouds. *Astrophys J* **823**: 27. DOI:[10.3847/0004-637X/823/1/27](https://doi.org/10.3847/0004-637X/823/1/27).
- Odstrčil D. 2003. Modeling 3-D solar wind structure. *Adv Space Res* **32**: 497–506. DOI:[10.1016/S0273-1177\(03\)00332-6](https://doi.org/10.1016/S0273-1177(03)00332-6).
- Odstrčil D, Pizzo VJ. 1999a. Three-dimensional propagation of CMEs in a structured solar wind flow: 1. CME launched within the streamer belt. *J Geophys Res* **104**: 483–492. DOI:[10.1029/1998JA900019](https://doi.org/10.1029/1998JA900019).
- Odstrčil D, Pizzo VJ. 1999b. Three-dimensional propagation of coronal mass ejections in a structured solar wind flow 2. CME launched adjacent to the streamer belt. *J Geophys Res* **104**: 493–504. DOI:[10.1029/1998JA900038](https://doi.org/10.1029/1998JA900038).
- Odstrčil D, Smith Z, Dryer M. 1996. Distortion of the heliospheric plasma sheet by interplanetary shocks. *Geophys Res Lett* **23**: 2521–2524. DOI:[10.1029/96GL00159](https://doi.org/10.1029/96GL00159).
- Odstrčil D, Riley P, Zhao XP. 2004. Numerical simulation of the 12 May 1997 interplanetary CME event. *J Geophys Res (Space Phys)* **109**: A02116. DOI:[10.1029/2003JA010135](https://doi.org/10.1029/2003JA010135).
- Richardson IG, Cane HV. 2010. Near-Earth interplanetary coronal mass ejections during solar cycle 23 (1996–2009): catalog and summary of properties. *Sol Phys* **264**: 189–237. DOI:[10.1007/s11207-010-9568-6](https://doi.org/10.1007/s11207-010-9568-6).
- Rostoker G. 1972. Geomagnetic indices. *Rev Geophys Space Phys* **10**: 935–950. DOI:[10.1029/RG010i004p00935](https://doi.org/10.1029/RG010i004p00935).
- Simunac KDC, Kistler LM, Galvin AB, Popecki MA, Farrugia CJ. 2009. In situ observations from STEREO/PLASTIC: a test for L5 space weather monitors. *Ann Geophys* **27**: 3805–3809. DOI:[10.5194/angeo-27-3805-2009](https://doi.org/10.5194/angeo-27-3805-2009). <https://www.ann-geophys.net/27/3805/2009/>
- Strugarek A, Janitzek N, Lee A, Löschl P, Seifert B, et al. 2015. A space weather mission concept: Observatories of the Solar Corona and Active Regions (OSCAR). *J Space Weather Space Clim* **5**: A4. DOI:[10.1051/swsc/2015003](https://doi.org/10.1051/swsc/2015003).
- Taktakishvili A, Kuznetsova M, MacNeice P, Hesse M, Rastätter L, Pulkkinen A, Chulaki A, Odstrčil D. 2009. Validation of the coronal mass ejection predictions at the Earth orbit estimated by ENLIL heliosphere cone model. *Space Weather* **7**: S03004. DOI:[10.1029/2008SW000448](https://doi.org/10.1029/2008SW000448).
- Temmer M, Reiss MA, Nikolic L, Hofmeister SJ, Veronig AM. 2017. Preconditioning of interplanetary space due to transient CME disturbances. *Astrophys J* **835**: 141. <http://stacks.iop.org/0004-637X/835/i=2/a=141>.
- Vourlidas A. 2015. Mission to the Sun-Earth L5 Lagrangian point: an optimal platform for space weather research. *Space Weather* **13**: 197–201. DOI:[10.1002/2015SW001173](https://doi.org/10.1002/2015SW001173).
- Vršnak B, Temmer M, Žic T, Taktakishvili A, Dumbović M, Möstl C, Veronig AM, Mays ML, Odstrčil D. 2014. Heliospheric propagation of coronal mass ejections: comparison of numerical WSA-ENLIL+Cone model and analytical drag-based model. *Astrophys J Suppl Ser* **213**: 21. DOI:[10.1088/0067-0049/213/2/21](https://doi.org/10.1088/0067-0049/213/2/21).
- Weinzierl M, Mackay DH, Yeates AR, Pevtsov AA. 2016. The possible impact of L5 magnetograms on non-potential solar coronal magnetic field simulations. *Astrophys J* **828**: 102. DOI:[10.3847/0004-637X/828/2/102](https://doi.org/10.3847/0004-637X/828/2/102).
- Wilks D, 2011, *Statistical methods in atmospheric sciences: an introduction*, Academic Press, Massachusetts, USA.
- Winslow RM. 2015. Interplanetary coronal mass ejections from MESSENGER orbital observations at Mercury. *J Geophys Res* **120**: 6101–6118. DOI:[10.1002/2015JA021200](https://doi.org/10.1002/2015JA021200).
- Xie H, Ofman L, Lawrence G. 2004. Cone model for halo CMEs: application to space weather forecasting. *J Geophys Res (Space Phys)* **109**: A03109. DOI:[10.1029/2003JA010226](https://doi.org/10.1029/2003JA010226).
- Zhao XP, Plunkett SP, Liu W. 2002. Determination of geometrical and kinematical properties of halo coronal mass ejections using the cone model. *J Geophys Res (Space Phys)* **107**: 1223. DOI:[10.1029/2001JA009143](https://doi.org/10.1029/2001JA009143).

Cite this article as: Wold AM, Mays ML, Taktakishvili A, Jian LK, Odstrčil D, MacNeice P. 2018. Verification of real-time WSA–ENLIL+Cone simulations of CME arrival-time at the CCMC from 2010 to 2016. *J. Space Weather Space Clim.* **8**: A17