Can One Predict Coronal Mass Ejection Arrival Times with Thirty-Minute Accuracy?

Gabor Toth[1], Bartholomeus van der Holst[1], and Ward Beecher Manchester IV[1]

[1]University of Michigan-Ann Arbor

February 27, 2023

Abstract

J. Schmidt and Cairns (2019) have recently claimed that they can predict Coronal Mass Ejection (CME) arrival times with an accuracy of 0.9+-1.9 hours for four separate events. They also stated that the accuracy gets better with increased grid resolution. Here, we show that combining their results with the Richardson extrapolation (Richardson and Gaunt, 1927), which is a standard technique in computational fluid dynamics, could predict the CME arrival time with 0.2+-0.26 hours accuracy. The CME arrival time errors of this model would lie in a 95% confidence interval [-0.21,0.61] h. We also show that the probability of getting these accurate arrival time predictions with a model with a standard deviation exceeding 2 hours is less than 0.1%, indicating that these results cannot be due to random chance. This unprecedented accuracy is about 20 times better than the current state-of-the-art prediction of CME arrival times with an average error of about +-10 hours. Based on our analysis there are only two possibilities: the results shown by J. Schmidt and Cairns (2019) were not obtained from reproducible numerical simulations, or their method combined by the Richardson extrapolation is in fact providing CME arrival times with half an hour accuracy. We believe that this latter interpretation is very unlikely to hold true. We also discuss how the peer-review process apparently failed to even question the validity of the results presented by Schmidt and Cairns (2019).

Can One Predict Coronal Mass Ejection Arrival Times with Thirty-Minute Accuracy?

Gábor Tóth, Bart van der Holst, Ward Manchester

Dept. of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA

Abstract

J. Schmidt and Cairns (2019) have recently claimed that they can predict Coronal Mass Ejection (CME) arrival times with an accuracy of 0.9±1.9 hours for four separate events. They also stated that the accuracy gets better with increased grid resolution. Here, we show that combining their results with the Richardson extrapolation (Richardson & Gaunt, 1927), which is a standard technique in computational fluid dynamics, could predict the CME arrival time with 0.2±0.26 hours accuracy. The CME arrival time errors of this model would lie in a 95% confidence interval $[-0.21, 0.61]$ h. We also show that the probability of getting these accurate arrival time predictions with a model with a standard deviation exceeding 2 hours is less than 0.1%, indicating that these results cannot be due to random chance. This unprecedented accuracy is about 20 times better than the current state-of-the-art prediction of CME arrival times with an average error of about ±10 hours. Based on our analysis there are only two possibilities: the results shown by J. Schmidt and Cairns (2019) were not obtained from reproducible numerical simulations, or their method combined by the Richardson extrapolation is in fact providing CME arrival times with half an hour accuracy. We believe that this latter interpretation is very unlikely to hold true. We also discuss how the peer-review process apparently failed to even question the validity of the results presented by J. Schmidt and Cairns (2019).

1 Introduction

Predicting the propagation of Coronal Mass Ejections (CMEs) and their arrival time at Earth has been a major goal of space weather prediction for decades. The ENLIL model (Odstrčil & Pizzo, 1999a, 1999b), for example, solves the ideal magnetohydrodynamic (MHD) equations from about 0.1 au $\approx 20R_s$ (solar radii) to the Earth orbit and beyond. For this model, the inner boundary conditions are provided by the Wang-Sheeley-Arge (WSA) model (Arge & Pizzo, 2000). CMEs are initiated with the empirical cone model based on flare observations and coronal white light images. Another approach is followed by the Alfvén Wave Solar atmosphere Model (AWSoM) (van der Holst et al., 2014) that is based on the BATS-R-US MHD code (Powell et al., 1999; Tóth et al., 2012), also widely used to model the solar corona, the heliosphere and the eruption and propagation of CMEs from the surface of the Sun (initiated by a flux rope model) to Earth and beyond (Tóth et al., 2007; Manchester et al., 2014; Jin et al., 2017, 2017). AWSoM solves the MHD equations extended with solar wind heating and acceleration due to Alfvén wave turbulence, radiative cooling and heat conduction. However, these first-principles models can only achieve about 10-hour accuracy predicting the CME arrival time (Wold et al., 2018, cf.). More recently, empirical and neural network based models were applied to this problem, but the typical error remains about ±10 hours (Riley et al., 2018; Amerstorfer et al., 2021, cf.).

J. Schmidt and Cairns (2019), heareafter SC, claim to have used an earlier coronal model based on BATS-R-US developed by Cohen et al. (2007), which relies on a spatially varying polytropic index derived from the Wang-Sheely-Arge (WSA) model (Arge & Pizzo, 2000) and achieved an unprecedented accuracy for predicting the CME arrival time: 0.9±1.9 hours. They describe their procedure of setting up the CME simulations using only information that is available prior to and within a few hours after the CME eruptions: the WSO magnetogram, the CME speed estimated from the CME Analysis Tool (CAT) using STEREO/LASCO C3 coronagraph images, and prior L1 *in situ* observations used for the WSA model and in turn for BATS-R-US. In addition, we have learned from the authors that the simulations were performed on a couple of CPU cores and they managed to run the model about three times faster than real time. This is worth contrasting with the computational resources used by ENLIL and AWSoM, which require 100s or even 1000s of CPU cores to run faster than real time.

SC have only published their work in form of a preprint on arxiv. An earlier version of the manuscript was submitted to the Geophysical Research Letters, where it was reviewed and rejected by one of us after a careful analysis of the output files requested and obtained from the authors. In spite of the highly critical review, Schmidt and Cairn have submitted the manuscript with a different title but essentially the same content to this journal, where it was actually accepted for publication. The only reason it was not published is that we contacted the editor regarding another manuscript with questionable content involving the same authors. In fact, these manuscripts are not outliers. As it is explained by SC, the "setup and analysis is refined from our earlier work simulating type II radio bursts and CMEs", which in fact resulted in four peer-reviewed and published works (J. M. Schmidt et al., 2013; J. M. Schmidt & Cairns, 2014, 2016; J. M. Schmidt et al., 2016). Therefore the content of SC can be safely considered to have similar quality and scientific value as these prior publications. It is therefore imperative to examine the validity of the results presented by SC.

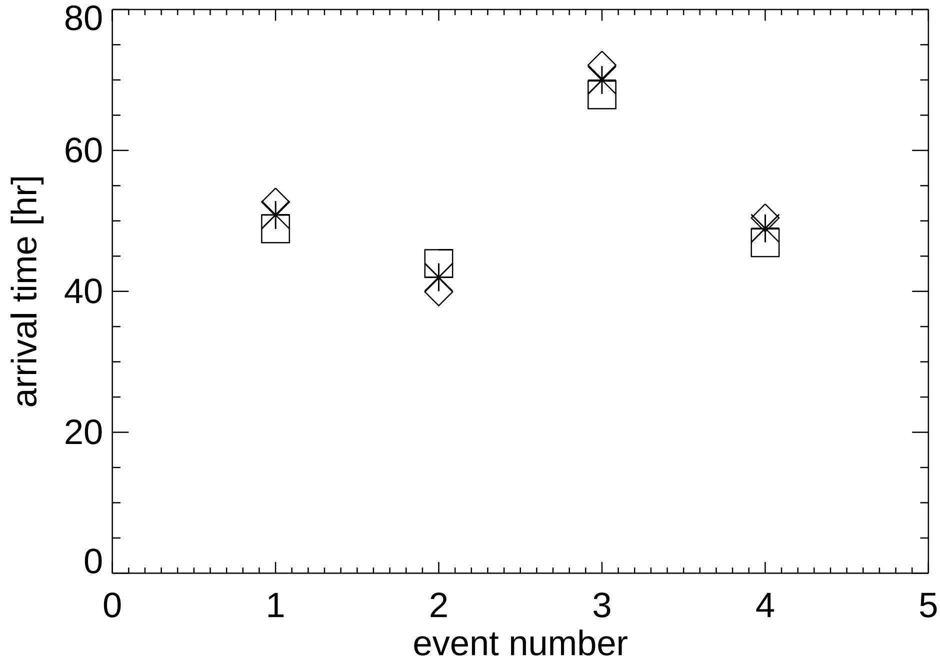

Figure 1. Observed and predicted arrival times at 1 au of four CME events (4 Sep 2017, 6 Sep 2017, 12 Feb 2018, and 29 Nov 2013 CME) recreated from Figure 4 in SC. The diamonds show observed arrival times, the squares and stars are simulation results at level 2 and level 5 grid refinements, respectively.

Looking at Figure 4 in SC, reproduced here as Figure 1, we have noticed that the distances between the observations (diamonds) and the model predictions obtained on two different computational grids (squares and stars) form a distinctive pattern: the distances between the three symbols appear to be approximately the same for all four events displayed. We show that if the figure showed the results of actual CME simulations, then this fact can be exploited to obtain an even more accurate estimate of the CME arrival time. Using the Richardson extrapolation (Richardson & Gaunt, 1927) the bias and standard deviation become 0.2 ± 0.26 hours, which is significantly better than the 0.9 ± 1.9 hours obtained by SC. We will also show that the agreement between observations and simulations cannot be attributed to luck. Since the four events happened in different years and/or have very different arrival times covering a wide range from about 40 hours to 72 hours, the technique must be applicable to most CMEs. This means that the model should provide extremely reliable and accurate information for operational space weather forecasters, which is important for our national security and human safety. Unfortunately, we cannot exclude the alternative explanation that the results shown by SC do not represent actual CME simulation results.

2 Predicting CME arrival times

To perform a quantitative evaluation of the results presented in Figure 4 of SC, we have digitized the figure and put the observed and simulated arrival times (relative to the eruption time) into Table 1. These values were also used to produce Figure 1 confirming that the values were extracted correctly.

Table 1. Simulated and observed CME arrival times for four events from Figure 4 in SC. The times are measured in hours from the eruption time. The error is the difference between the observed and simulated times.

ID	Date	Observed	Model1	Model2	Error1	Error2	Error1/Error2
1	Sep 04, 2017	52.68	48.87	50.85	3.80	1.83	2.08
2	Sep 06, 2017	39.95	43.94	42.00	−4.01	−2.07	1.94
3	Feb 12, 2018	72.11	67.89	69.98	4.23	2.13	1.98
4	Nov 29, 2013	50.42	46.90	48.94	3.52	1.48	2.38
Average magnitude					3.89	1.87	2.09

The errors, Error1 and Error2 of the two models Model1 and Model2, corresponding to Refinement Level 2 and 5 in SC, are remarkably constant across the four events, and the ratio of the errors is approximately 2.1. We note that SC does not define what refinement levels 2 and 5 actually mean, so we simply assume here that the model with refinement level 5 is more accurate than the one with level 2 due to better grid resolution. This allows us to use the idea of the Richardson extrapolation, which improves the numerical accuracy by estimating the exact solution from numerical solutions at two different grid resolutions. The leading term of numerical error can be written as

$$E(\Delta x) = T_{\text{exact}} - T(\Delta x) = K\Delta x^n + O(\Delta x^{n+1}) \tag{1}$$

where T_{exact} is the exact (observed) arrival time, $T(\Delta x)$ is the arrival time obtained by a simulation using grid cell size Δx, K is some problem (but not grid) dependent constant coefficient, n is the order of the scheme and $O(\Delta x^{n+1})$ are contributions from higher order terms. For a first order accurate scheme, which is appropriate for shock propagation, $n = 1$, so the leading error term is proportional to the grid resolution. Equation (1) can be solved for T_{exact} if $T(\Delta x)$ is known for at least two different grid resolutions:

$$T_{\text{exact}} = 2T(\Delta x) - T(2\Delta x) + O(\Delta x^2) \tag{2}$$

We define the Richardson extrapolated arrival time as

$$T_R = 2T_2 - T_1 \tag{3}$$

where T_1 and T_2 are the arrival times predicted by models 1 and 2 using grid resolutions differing by a factor of 2. T_R has a much improved accuracy compared to the accuracy of the original simulation results T_1 and T_2.

Table 2. Observed and extrapolated CME arrival times for four events. The times are measured in hours from the eruption time. The last column is the absolute value of the error.

ID i	Date	Observed T_i	Extrapolated $T_{i,R}$	Error $T_{i,R} - T_i$
1	Sep 04, 2017	52.68	52.82	0.14
2	Sep 06, 2017	39.93	40.06	0.13
3	Feb 12, 2018	72.11	72.07	-0.04
4	Nov 29, 2013	50.42	50.99	0.57
Mean absolute error				0.22
Mean ± one standard deviation				0.2 ± 0.26

3 Statistical Analysis and Probability Estimates

Table 2 shows that the mean absolute error of the extrapolated arrival time is about 0.218 hours, which is useful information, but not suitable for statistical analysis. To better quantify the performance of the new model, we calculate an unbiased estimate and a 95% confidence interval for the arrival time errors.

The sample size is $N = 4$. The average of the errors, the bias, is

$$B = \frac{1}{N} \sum_{i=1}^{N} (T_{i,R} - T_i) = 0.2\,h \tag{4}$$

and the standard deviation S is

$$S = \sqrt{\frac{\sum_{i=1}^{N} (T_{i,R} - T_i - B)^2}{N-1}} = 0.26\,\mathrm{h} \tag{5}$$

where T_i is the observed arrival time for event i and $T_{i,R}$ is the Richardson extrapolated time calculated from Equation 3. The 95% confidence interval for the error $T_R - T$ is $B \pm tS/\sqrt{N}$, where $t = 3.182$ from the T-distribution for $p = 0.025$ and $N - 1 = 3$ degrees of freedom:

$$(T_R - T) \in [-0.21, 0.61]\,\mathrm{h} \tag{6}$$

We conclude that there is a 95% chance that the model will produce arrival time predictions with errors less than 37 minutes, while the average error is only 12 minutes.

Finally, it is important to check if the small errors in Table 2 are statistically significant, or they can be attributed to simple luck. Let us assume that the new model with the extrapolation has no bias, $\mu = 0$, and its standard deviation is $\sigma = 2\,\mathrm{h}$. The quantity

$$X^2 = \frac{\sum_{i=1}^{N} (T_{i,R} - T_i)^2}{\sigma^2} = 0.089 \tag{7}$$

follows the $\chi^2(N, p)$ distribution since the mean value is assumed to be known. For $N = 4$, we find that there is only $p = 0.1\%$ chance that $X^2 \leq 0.089$ by pure luck. If σ was larger than 2 hours, this probability would be even less. We can safely conclude that the model is indeed capable of predicting the CME arrival time with high accuracy, even higher than the original SC model, assuming that the SC model results are true.

4 On the Validity of the CME Simulations Presented by SC

In addition to the improbable accuracy of the CME arrival time predictions, there are a number of inexplicable inconsistencies in SC, which raise grave concern over the validity and reporting of their CME simulations. First, the flux rope electric current was increased by a factor of ten for a more refined spatial grid. In fact, the opposite should be true. Reduced numerical diffusion brought with a refined grid should allow the model to produce the same CME speed with a *reduced* electric current. Second, the magnitude of the electric currents shown in Figure 1 is more than an order of magnitude too large when compared to previously simulated results. (Manchester et al., 2012) used the TD flux rope and obtained CME speeds of 800 and 1000 km/s respectively with currents of 2.5×10^{11} and 3.25×10^{11} A respectively. Similarly, the Halloween event CME (Tóth et al., 2007; Manchester et al., 2008) was driven with a current of 6×10^{11} A. Currents of $10^{12}-10^{13}$ A would produce extraordinarily fast CMEs with speeds exceeding 3000 km/s, far beyond what is described by SC. Third, the interplanetary magnetic field strengths shown in Figure 2 of SC are an order of magnitude too strong, $100-400$ nT near Earth. These results are entirely unphysical and inconsistent with the field strengths shown in Figure 3 of SC where we find $B_z \approx 15$ nT and nearly constant, in sharp contradiction with the magnitude and significant spatial structure in their Figure 2. Finally, there is no possible explanation for how the simulated CME events on September 7 cannot reach the Earth, when the Earth is directly in front of their path.

5 Conclusions

In this paper, we have examined the work of SC, who claimed to predict CME arrival times with 0.9 ± 1.9 h accuracy. Using the standard Richardson extrapolation technique, we have further improved the accuracy of the SC model to an average prediction time error of 0.2 ± 0.26 hours. We showed that it is practically impossible that the good agreement between observations and simulation results obtained by SC was simply a lucky coincidence. The likelihood that an MHD model can be used to predict CME arrival times with 30-minute accuracy is exceedingly small, especially with no model enhancements to explain the more than an order of magnitude improvement over prior work using the same model. This result, unfortunately, leaves only one reasonable explanation for the SC results: they were most likely not obtained by reproducible numerical simulations. The content of prior publications (J. M. Schmidt et al., 2013; J. M. Schmidt & Cairns, 2014, 2016; J. M. Schmidt et al., 2016) that according to SC used the same "technique" are similarly questionable.

It appears that the peer review process worked when the original manuscript was submitted to the Geophysical Research Letters, but it failed when the same manuscript (with a different title) was submitted to this journal. It also seems likely that several published papers (J. M. Schmidt et al., 2013; J. M. Schmidt & Cairns, 2014, 2016; J. M. Schmidt et al., 2016) with questionable content have slipped through the peer review process. Reviewers cannot be experts in everything, but choosing reviewers with the right expertise can reduce the chances of such incidents. Tracking submitted and rejected manuscripts in a data base shared by several journals could be another safeguard. Most importantly, the requirements of reproducibility, open data and open software for published work should improve the reliability of the published scientific content dramatically. In particular, the invalidity of the SC results was abundantly apparent for the reviewer who received their input and output files. Readers and reviewers who only rely on the manuscript and published papers may or may not be able to distinguish genuine science from the type of content presented by SC.

6 Open Research

All data used in this paper are contained in Table 1. The Space Weather Modeling Framework including (BATS-R-US/AWSoM) is an open-source code available at https://github.com/MSTEM-QUDA with a full version history.

References

Amerstorfer, T., Hinterreiter, J., Reiss, M. A., Möstl, C., Davies, J. A., Bailey, R. L., ... Harrison, R. A. (2021). Evaluation of CME arrival prediction using ensemble modeling based on heliospheric imaging observations. *Space Weather*, *19*, e2020SW002553. doi: 10.1029/2020SW002553

Arge, C., & Pizzo, V. (2000). Improvement in the prediction of solar wind conditions using near-real time solar magnetic field updates. *J. Geophys. Res.*, *105*, 10465-10479. doi: 10.1029/1999JA000262

Cohen, O., Sokolov, I., Roussev, I., Arge, C., Manchester, W., Gombosi, T., ... Velli, M. (2007). A semi-empirical magnetohydrodynamical model of the solar wind. *Astrophys. J.*, *654*, L163-L166. doi: 10.1086/511154

Jin, M., Manchester, W. B., van der Holst, B., Sokolov, I., Tóth, G., Mullinix, R. E., ... Gombosi, T. I. (2017, January). Data-constrained Coronal Mass Ejections in a Global Magnetohydrodynamics Model. *Astrophys. J.*, *834*(2), 173. doi: 10.3847/1538-4357/834/2/173

Jin, M., Manchester, W. B., van der Holst, B., Sokolov, I., Tóth, G., Vourlidas, A., ... Gombosi, T. I. (2017, January). Chromosphere to 1 AU Simulation of the 2011 March 7th Event: A Comprehensive Study of Coronal Mass Ejection Propagation. *Astrophys. J.*, *834*(2), 172. doi: 10.3847/1538-4357/834/2/172

Manchester, I., W. B., van der Holst, B., & Lavraud, B. (2014, June). Flux rope evolution in interplanetary coronal mass ejections: the 13 May 2005 event. *Plasma Physics and Controlled Fusion*, *56*(6), 064006. doi: 10.1088/0741-3335/56/6/064006

Manchester, W. B., van der Holst, B., Tóth, G., & Gombosi, . T. I. (2012). The coupled evolution of electrons and ions in coronal mass ejection-driven shocks. *Astrophys. J.*, *756*. doi: 10.1088/0004-637X/756/1/81

Manchester, W. B., Vourlidas, A., Tóth, G., Lugaz, N., Roussev, I. I., Sokolov, I. V., ... Opher, M. (2008). Three-dimensional MHD simulation of the 2003 October 28 coronal mass ejection: Comparison with LASCO coronagraph observations. *Astrophys. J.*, *684*, 1448-1460. doi: 10.1086/590231

Odstrčil, D., & Pizzo, V. J. (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

Odstrčil, D., & Pizzo, V. J. (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*, 493–503. doi: 10.1029/1998JA900038

Powell, K., Roe, P., Linde, T., Gombosi, T., & De Zeeuw, D. L. (1999). A solution-adaptive upwind scheme for ideal magnetohydrodynamics. *J. Comput. Phys.*, *154*, 284-309. doi: 10.1006/jcph.1999.6299

Richardson, L. F., & Gaunt, J. A. (1927). The deferred approach to the limit. *Philos. Trans. of the Royal Soc.*, *226*, 299. doi: 10.1098/rsta.1927.0008

Riley, P., Mays, M. L., Andries, J., Amerstorfer, T., Biesecker, D., Delouille, V., ... Zhao, X. (2018). Forecasting the arrival time of coronal mass ejections: Analysis of the ccmc cme scoreboard. *Space Weather*, *16*(9), 1245-1260. doi: 10.1029/2018SW001962

Schmidt, J., & Cairns, I. (2019). Hit or miss, arrival time, and b_z orientation predictions of bats-r-us cme simulations at 1 au. *arXiv*, *1905*, e08961. doi: 10.48550/arXiv.1905.08961

Schmidt, J. M., & Cairns, I. H. (2014). Type ii solar radio bursts predicted by 3d mhd cme and kinetic radio emission simulations. *J. Geophys. Res.*, *119*, 69. doi: 10.1002/2013JA019349

Schmidt, J. M., & Cairns, I. H. (2016). Quantitative prediction of type ii solar radio emission from the sun to 1 au. *Geophys. Res. Lett.*, *43*, 50. doi: 10.1002/2015GL067271

Schmidt, J. M., Cairns, I. H., Cyr, O. C. S., Xie, H., & Gopalswamy, N. (2016). Cme flux rope and shock identifications and locations: Comparison of white light data, graduated cylindrical shell (gcs) model, and mhd simulations. *J. Geophys. Res.*, *121*. doi: 10.1002/2015JA021805

Schmidt, J. M., Cairns, I. H., & Hillan, D. S. (2013). *Astrophys. J.*, *773*, L30. doi: 10.1088/2041-8205/773/2/L30

Tóth, G., de Zeeuw, D. L., Gombosi, T. I., Manchester, W. B., Ridley, A. J., Sokolov, I. V., & Roussev, I. I. (2007, June). Sun-to-thermosphere simulation of the 28-30 October 2003 storm with the Space Weather Modeling Framework. *Space Weather*, *5*(6), 06003. doi: 10.1029/2006SW000272

Tóth, G., van der Holst, B., Sokolov, I. V., Zeeuw, D. L. D., Gombosi, T. I., Fang, F., ... Opher, M. (2012). Adaptive numerical algorithms in space weather modeling. *J. Comput. Phys.*, *231*, 870–903. doi: 10.1016/j.jcp.2011.02.006

van der Holst, B., Sokolov, I., Meng, X., Jin, M., Manchester, W. B., Tóth, G., & Gombosi, T. I. (2014). Alfvén wave solar model (AWSOM): Coronal heating. *Astrophys. J.*, *782*, 81. doi: 10.1088/0004-637X/782/2/81

Wold, A. M., Mays, M. L., Taktakishvili, A., Jian, L. K., Odstrcil, 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. doi: 10.1051/swsc/2018005

Can One Predict Coronal Mass Ejection Arrival Times with Thirty-Minute Accuracy?

Gábor Tóth, Bart van der Holst, Ward Manchester

Dept. of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA

Abstract

J. Schmidt and Cairns (2019) have recently claimed that they can predict Coronal Mass Ejection (CME) arrival times with an accuracy of 0.9±1.9 hours for four separate events. They also stated that the accuracy gets better with increased grid resolution. Here, we show that combining their results with the Richardson extrapolation (Richardson & Gaunt, 1927), which is a standard technique in computational fluid dynamics, could predict the CME arrival time with 0.2±0.26 hours accuracy. The CME arrival time errors of this model would lie in a 95% confidence interval $[-0.21, 0.61]$ h. We also show that the probability of getting these accurate arrival time predictions with a model with a standard deviation exceeding 2 hours is less than 0.1%, indicating that these results cannot be due to random chance. This unprecedented accuracy is about 20 times better than the current state-of-the-art prediction of CME arrival times with an average error of about ±10 hours. Based on our analysis there are only two possibilities: the results shown by J. Schmidt and Cairns (2019) were not obtained from reproducible numerical simulations, or their method combined by the Richardson extrapolation is in fact providing CME arrival times with half an hour accuracy. We believe that this latter interpretation is very unlikely to hold true. We also discuss how the peer-review process apparently failed to even question the validity of the results presented by J. Schmidt and Cairns (2019).

1 Introduction

Predicting the propagation of Coronal Mass Ejections (CMEs) and their arrival time at Earth has been a major goal of space weather prediction for decades. The ENLIL model (Odstrčil & Pizzo, 1999a, 1999b), for example, solves the ideal magnetohydrodynamic (MHD) equations from about 0.1 au $\approx 20R_s$ (solar radii) to the Earth orbit and beyond. For this model, the inner boundary conditions are provided by the Wang-Sheeley-Arge (WSA) model (Arge & Pizzo, 2000). CMEs are initiated with the empirical cone model based on flare observations and coronal white light images. Another approach is followed by the Alfvén Wave Solar atmosphere Model (AWSoM) (van der Holst et al., 2014) that is based on the BATS-R-US MHD code (Powell et al., 1999; Tóth et al., 2012), also widely used to model the solar corona, the heliosphere and the eruption and propagation of CMEs from the surface of the Sun (initiated by a flux rope model) to Earth and beyond (Tóth et al., 2007; Manchester et al., 2014; Jin et al., 2017, 2017). AWSoM solves the MHD equations extended with solar wind heating and acceleration due to Alfvén wave turbulence, radiative cooling and heat conduction. However, these first-principles models can only achieve about 10-hour accuracy predicting the CME arrival time (Wold et al., 2018, cf.). More recently, empirical and neural network based models were applied to this problem, but the typical error remains about ±10 hours (Riley et al., 2018; Amerstorfer et al., 2021, cf.).

J. Schmidt and Cairns (2019), heareafter SC, claim to have used an earlier coronal model based on BATS-R-US developed by Cohen et al. (2007), which relies on a spatially varying polytropic index derived from the Wang-Sheely-Arge (WSA) model (Arge & Pizzo, 2000) and achieved an unprecedented accuracy for predicting the CME arrival time: 0.9±1.9 hours. They describe their procedure of setting up the CME simulations using only information that is available prior to and within a few hours after the CME eruptions: the WSO magnetogram, the CME speed estimated from the CME Analysis Tool (CAT) using STEREO/LASCO C3 coronagraph images, and prior L1 *in situ* observations used for the WSA model and in turn for BATS-R-US. In addition, we have learned from the authors that the simulations were performed on a couple of CPU cores and they managed to run the model about three times faster than real time. This is worth contrasting with the computational resources used by ENLIL and AWSoM, which require 100s or even 1000s of CPU cores to run faster than real time.

SC have only published their work in form of a preprint on arxiv. An earlier version of the manuscript was submitted to the Geophysical Research Letters, where it was reviewed and rejected by one of us after a careful analysis of the output files requested and obtained from the authors. In spite of the highly critical review, Schmidt and Cairn have submitted the manuscript with a different title but essentially the same content to this journal, where it was actually accepted for publication. The only reason it was not published is that we contacted the editor regarding another manuscript with questionable content involving the same authors. In fact, these manuscripts are not outliers. As it is explained by SC, the "setup and analysis is refined from our earlier work simulating type II radio bursts and CMEs", which in fact resulted in four peer-reviewed and published works (J. M. Schmidt et al., 2013; J. M. Schmidt & Cairns, 2014, 2016; J. M. Schmidt et al., 2016). Therefore the content of SC can be safely considered to have similar quality and scientific value as these prior publications. It is therefore imperative to examine the validity of the results presented by SC.

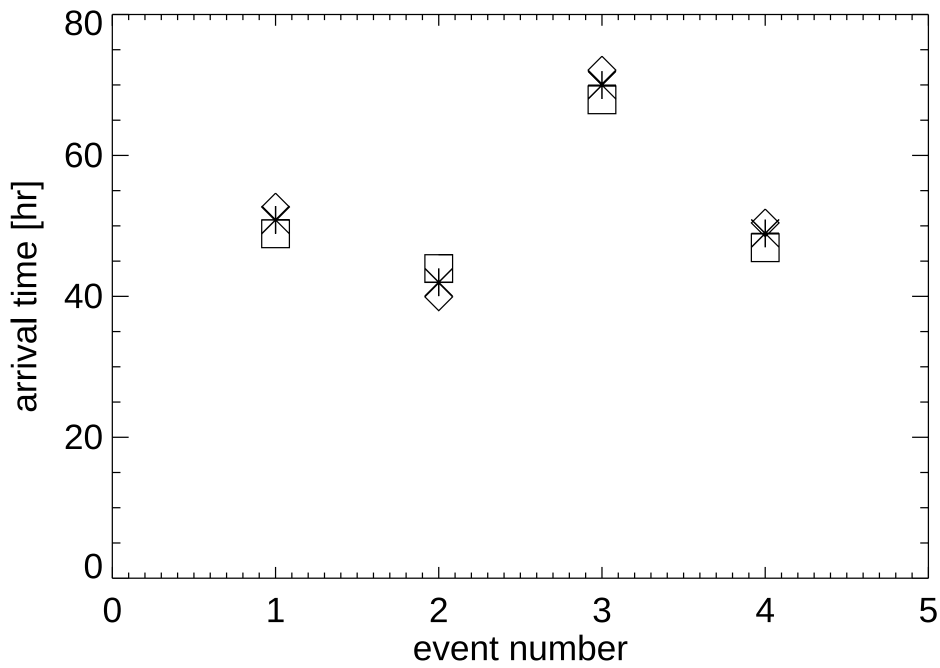

Figure 1. Observed and predicted arrival times at 1 au of four CME events (4 Sep 2017, 6 Sep 2017, 12 Feb 2018, and 29 Nov 2013 CME) recreated from Figure 4 in SC. The diamonds show observed arrival times, the squares and stars are simulation results at level 2 and level 5 grid refinements, respectively.

Looking at Figure 4 in SC, reproduced here as Figure 1, we have noticed that the distances between the observations (diamonds) and the model predictions obtained on two different computational grids (squares and stars) form a distinctive pattern: the distances between the three symbols appear to be approximately the same for all four events displayed. We show that if the figure showed the results of actual CME simulations, then this fact can be exploited to obtain an even more accurate estimate of the CME arrival time. Using the Richardson extrapolation (Richardson & Gaunt, 1927) the bias and standard deviation become 0.2 ± 0.26 hours, which is significantly better than the 0.9 ± 1.9 hours obtained by SC. We will also show that the agreement between observations and simulations cannot be attributed to luck. Since the four events happened in different years and/or have very different arrival times covering a wide range from about 40 hours to 72 hours, the technique must be applicable to most CMEs. This means that the model should provide extremely reliable and accurate information for operational space weather forecasters, which is important for our national security and human safety. Unfortunately, we cannot exclude the alternative explanation that the results shown by SC do not represent actual CME simulation results.

2 Predicting CME arrival times

To perform a quantitative evaluation of the results presented in Figure 4 of SC, we have digitized the figure and put the observed and simulated arrival times (relative to the eruption time) into Table 1. These values were also used to produce Figure 1 confirming that the values were extracted correctly.

Table 1. Simulated and observed CME arrival times for four events from Figure 4 in SC. The times are measured in hours from the eruption time. The error is the difference between the observed and simulated times.

ID	Date	Observed	Model1	Model2	Error1	Error2	Error1/Error2
1	Sep 04, 2017	52.68	48.87	50.85	3.80	1.83	2.08
2	Sep 06, 2017	39.95	43.94	42.00	−4.01	−2.07	1.94
3	Feb 12, 2018	72.11	67.89	69.98	4.23	2.13	1.98
4	Nov 29, 2013	50.42	46.90	48.94	3.52	1.48	2.38
Average magnitude					3.89	1.87	2.09

The errors, Error1 and Error2 of the two models Model1 and Model2, corresponding to Refinement Level 2 and 5 in SC, are remarkably constant across the four events, and the ratio of the errors is approximately 2.1. We note that SC does not define what refinement levels 2 and 5 actually mean, so we simply assume here that the model with refinement level 5 is more accurate than the one with level 2 due to better grid resolution. This allows us to use the idea of the Richardson extrapolation, which improves the numerical accuracy by estimating the exact solution from numerical solutions at two different grid resolutions. The leading term of numerical error can be written as

$$E(\Delta x) = T_{\text{exact}} - T(\Delta x) = K \Delta x^n + O(\Delta x^{n+1}) \tag{1}$$

where T_{exact} is the exact (observed) arrival time, $T(\Delta x)$ is the arrival time obtained by a simulation using grid cell size Δx, K is some problem (but not grid) dependent constant coefficient, n is the order of the scheme and $O(\Delta x^{n+1})$ are contributions from higher order terms. For a first order accurate scheme, which is appropriate for shock propagation, $n = 1$, so the leading error term is proportional to the grid resolution. Equation (1) can be solved for T_{exact} if $T(\Delta x)$ is known for at least two different grid resolutions:

$$T_{\text{exact}} = 2T(\Delta x) - T(2\Delta x) + O(\Delta x^2) \tag{2}$$

We define the Richardson extrapolated arrival time as

$$T_R = 2T_2 - T_1 \tag{3}$$

where T_1 and T_2 are the arrival times predicted by models 1 and 2 using grid resolutions differing by a factor of 2. T_R has a much improved accuracy compared to the accuracy of the original simulation results T_1 and T_2.

Table 2. Observed and extrapolated CME arrival times for four events. The times are measured in hours from the eruption time. The last column is the absolute value of the error.

ID i	Date	Observed T_i	Extrapolated $T_{i,R}$	Error $T_{i,R} - T_i$
1	Sep 04, 2017	52.68	52.82	0.14
2	Sep 06, 2017	39.93	40.06	0.13
3	Feb 12, 2018	72.11	72.07	-0.04
4	Nov 29, 2013	50.42	50.99	0.57
Mean absolute error				0.22
Mean $\pm$ one standard deviation				0.2 ± 0.26

3 Statistical Analysis and Probability Estimates

Table 2 shows that the mean absolute error of the extrapolated arrival time is about 0.218 hours, which is useful information, but not suitable for statistical analysis. To better quantify the performance of the new model, we calculate an unbiased estimate and a 95% confidence interval for the arrival time errors.

The sample size is $N = 4$. The average of the errors, the bias, is

$$B = \frac{1}{N} \sum_{i=1}^{N} (T_{i,R} - T_i) = 0.2\, h \tag{4}$$

and the standard deviation S is

$$S = \sqrt{\frac{\sum_{i=1}^{N} (T_{i,R} - T_i - B)^2}{N-1}} = 0.26\,\mathrm{h} \tag{5}$$

where T_i is the observed arrival time for event i and $T_{i,R}$ is the Richardson extrapolated time calculated from Equation 3. The 95% confidence interval for the error $T_R - T$ is $B \pm tS/\sqrt{N}$, where $t = 3.182$ from the T-distribution for $p = 0.025$ and $N - 1 = 3$ degrees of freedom:

$$(T_R - T) \in [-0.21, 0.61]\,\mathrm{h} \tag{6}$$

We conclude that there is a 95% chance that the model will produce arrival time predictions with errors less than 37 minutes, while the average error is only 12 minutes.

Finally, it is important to check if the small errors in Table 2 are statistically significant, or they can be attributed to simple luck. Let us assume that the new model with the extrapolation has no bias, $\mu = 0$, and its standard deviation is $\sigma = 2\,\mathrm{h}$. The quantity

$$X^2 = \frac{\sum_{i=1}^{N} (T_{i,R} - T_i)^2}{\sigma^2} = 0.089 \tag{7}$$

follows the $\chi^2(N, p)$ distribution since the mean value is assumed to be known. For $N = 4$, we find that there is only $p = 0.1\%$ chance that $X^2 \leq 0.089$ by pure luck. If σ was larger than 2 hours, this probability would be even less. We can safely conclude that the model is indeed capable of predicting the CME arrival time with high accuracy, even higher than the original SC model, assuming that the SC model results are true.

4 On the Validity of the CME Simulations Presented by SC

In addition to the improbable accuracy of the CME arrival time predictions, there are a number of inexplicable inconsistencies in SC, which raise grave concern over the validity and reporting of their CME simulations. First, the flux rope electric current was increased by a factor of ten for a more refined spatial grid. In fact, the opposite should be true. Reduced numerical diffusion brought with a refined grid should allow the model to produce the same CME speed with a *reduced* electric current. Second, the magnitude of the electric currents shown in Figure 1 is more than an order of magnitude too large when compared to previously simulated results. (Manchester et al., 2012) used the TD flux rope and obtained CME speeds of 800 and 1000 km/s respectively with currents of 2.5×10^{11} and 3.25×10^{11} A respectively. Similarly, the Halloween event CME (Tóth et al., 2007; Manchester et al., 2008) was driven with a current of 6×10^{11} A. Currents of $10^{12} - 10^{13}$ A would produce extraordinarily fast CMEs with speeds exceeding 3000 km/s, far beyond what is described by SC. Third, the interplanetary magnetic field strengths shown in Figure 2 of SC are an order of magnitude too strong, $100 - 400$ nT near Earth. These results are entirely unphysical and inconsistent with the field strengths shown in Figure 3 of SC where we find $B_z \approx 15$ nT and nearly constant, in sharp contradiction with the magnitude and significant spatial structure in their Figure 2. Finally, there is no possible explanation for how the simulated CME events on September 7 cannot reach the Earth, when the Earth is directly in front of their path.

5 Conclusions

In this paper, we have examined the work of SC, who claimed to predict CME arrival times with 0.9 ± 1.9 h accuracy. Using the standard Richardson extrapolation technique, we have further improved the accuracy of the SC model to an average prediction time error of 0.2 ± 0.26 hours. We showed that it is practically impossible that the good agreement between observations and simulation results obtained by SC was simply a lucky coincidence. The likelihood that an MHD model can be used to predict CME arrival times with 30-minute accuracy is exceedingly small, especially with no model enhancements to explain the more than an order of magnitude improvement over prior work using the same model. This result, unfortunately, leaves only one reasonable explanation for the SC results: they were most likely not obtained by reproducible numerical simulations. The content of prior publications (J. M. Schmidt et al., 2013; J. M. Schmidt & Cairns, 2014, 2016; J. M. Schmidt et al., 2016) that according to SC used the same "technique" are similarly questionable.

It appears that the peer review process worked when the original manuscript was submitted to the Geophysical Research Letters, but it failed when the same manuscript (with a different title) was submitted to this journal. It also seems likely that several published papers (J. M. Schmidt et al., 2013; J. M. Schmidt & Cairns, 2014, 2016; J. M. Schmidt et al., 2016) with questionable content have slipped through the peer review process. Reviewers cannot be experts in everything, but choosing reviewers with the right expertise can reduce the chances of such incidents. Tracking submitted and rejected manuscripts in a data base shared by several journals could be another safeguard. Most importantly, the requirements of reproducibility, open data and open software for published work should improve the reliability of the published scientific content dramatically. In particular, the invalidity of the SC results was abundantly apparent for the reviewer who received their input and output files. Readers and reviewers who only rely on the manuscript and published papers may or may not be able to distinguish genuine science from the type of content presented by SC.

6 Open Research

All data used in this paper are contained in Table 1. The Space Weather Modeling Framework including (BATS-R-US/AWSoM) is an open-source code available at https://github.com/MSTEM-QUDA with a full version history.

References

Amerstorfer, T., Hinterreiter, J., Reiss, M. A., Möstl, C., Davies, J. A., Bailey, R. L., … Harrison, R. A. (2021). Evaluation of CME arrival prediction using ensemble modeling based on heliospheric imaging observations. *Space Weather*, *19*, e2020SW002553. doi: 10.1029/2020SW002553

Arge, C., & Pizzo, V. (2000). Improvement in the prediction of solar wind conditions using near-real time solar magnetic field updates. *J. Geophys. Res.*, *105*, 10465-10479. doi: 10.1029/1999JA000262

Cohen, O., Sokolov, I., Roussev, I., Arge, C., Manchester, W., Gombosi, T., … Velli, M. (2007). A semi-empirical magnetohydrodynamical model of the solar wind. *Astrophys. J.*, *654*, L163-L166. doi: 10.1086/511154

Jin, M., Manchester, W. B., van der Holst, B., Sokolov, I., Tóth, G., Mullinix, R. E., … Gombosi, T. I. (2017, January). Data-constrained Coronal Mass Ejections in a Global Magnetohydrodynamics Model. *Astrophys. J.*, *834*(2), 173. doi: 10.3847/1538-4357/834/2/173

Jin, M., Manchester, W. B., van der Holst, B., Sokolov, I., Tóth, G., Vourlidas, A., … Gombosi, T. I. (2017, January). Chromosphere to 1 AU Simulation of the 2011 March 7th Event: A Comprehensive Study of Coronal Mass Ejection Propagation. *Astrophys. J.*, *834*(2), 172. doi: 10.3847/1538-4357/834/2/172

Manchester, I., W. B., van der Holst, B., & Lavraud, B. (2014, June). Flux rope evolution in interplanetary coronal mass ejections: the 13 May 2005 event. *Plasma Physics and Controlled Fusion*, *56*(6), 064006. doi: 10.1088/0741-3335/56/6/064006

Manchester, W. B., van der Holst, B., Tóth, G., & Gombosi, . T. I. (2012). The coupled evolution of electrons and ions in coronal mass ejection-driven shocks. *Astrophys. J.*, *756*. doi: 10.1088/0004-637X/756/1/81

Manchester, W. B., Vourlidas, A., Tóth, G., Lugaz, N., Roussev, I. I., Sokolov, I. V., … Opher, M. (2008). Three-dimensional MHD simulation of the 2003 October 28 coronal mass ejection: Comparison with LASCO coronagraph observations. *Astrophys. J.*, *684*, 1448-1460. doi: 10.1086/590231

Odstrčil, D., & Pizzo, V. J. (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

Odstrčil, D., & Pizzo, V. J. (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*, 493–503. doi: 10.1029/1998JA900038

Powell, K., Roe, P., Linde, T., Gombosi, T., & De Zeeuw, D. L. (1999). A solution-adaptive upwind scheme for ideal magnetohydrodynamics. *J. Comput. Phys.*, *154*, 284-309. doi: 10.1006/jcph.1999.6299

Richardson, L. F., & Gaunt, J. A. (1927). The deferred approach to the limit. *Philos. Trans. of the Royal Soc.*, *226*, 299. doi: 10.1098/rsta.1927.0008

Riley, P., Mays, M. L., Andries, J., Amerstorfer, T., Biesecker, D., Delouille, V., … Zhao, X. (2018). Forecasting the arrival time of coronal mass ejections: Analysis of the ccmc cme scoreboard. *Space Weather*, *16*(9), 1245-1260. doi: 10.1029/2018SW001962

Schmidt, J., & Cairns, I. (2019). Hit or miss, arrival time, and b_z orientation predictions of bats-r-us cme simulations at 1 au. *arXiv*, *1905*, e08961. doi: 10.48550/arXiv.1905.08961

Schmidt, J. M., & Cairns, I. H. (2014). Type ii solar radio bursts predicted by 3d mhd cme and kinetic radio emission simulations. *J. Geophys. Res.*, *119*, 69. doi: 10.1002/2013JA019349

Schmidt, J. M., & Cairns, I. H. (2016). Quantitative prediction of type ii solar radio emission from the sun to 1 au. *Geophys. Res. Lett.*, *43*, 50. doi: 10.1002/2015GL067271

Schmidt, J. M., Cairns, I. H., Cyr, O. C. S., Xie, H., & Gopalswamy, N. (2016). Cme flux rope and shock identifications and locations: Comparison of white light data, graduated cylindrical shell (gcs) model, and mhd simulations. *J. Geophys. Res.*, *121*. doi: 10.1002/2015JA021805

Schmidt, J. M., Cairns, I. H., & Hillan, D. S. (2013). *Astrophys. J.*, *773*, L30. doi: 10.1088/2041-8205/773/2/L30

Tóth, G., de Zeeuw, D. L., Gombosi, T. I., Manchester, W. B., Ridley, A. J., Sokolov, I. V., & Roussev, I. I. (2007, June). Sun-to-thermosphere simulation of the 28-30 October 2003 storm with the Space Weather Modeling Framework. *Space Weather*, *5*(6), 06003. doi: 10.1029/2006SW000272

Tóth, G., van der Holst, B., Sokolov, I. V., Zeeuw, D. L. D., Gombosi, T. I., Fang, F., ... Opher, M. (2012). Adaptive numerical algorithms in space weather modeling. *J. Comput. Phys.*, *231*, 870–903. doi: 10.1016/j.jcp.2011.02.006

van der Holst, B., Sokolov, I., Meng, X., Jin, M., Manchester, W. B., Tóth, G., & Gombosi, T. I. (2014). Alfvén wave solar model (AWSOM): Coronal heating. *Astrophys. J.*, *782*, 81. doi: 10.1088/0004-637X/782/2/81

Wold, A. M., Mays, M. L., Taktakishvili, A., Jian, L. K., Odstrcil, 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. doi: 10.1051/swsc/2018005