Version 2018/11/10 00:49
Preprint typeset using LaTeX style emulateapj v. 01/23/15

NASA/NOAA MOU ANNEX YEAR 1 REPORT: EVALUATING MODEL ADVANCEMENTS FOR PREDICTING CME ARRIVAL TIME

CCMC: M. L. Mays[1], P. J. MacNeice[1], A. Taktakishvili[2,1], J. Merka[3,1], C. Wiegand[1],
SWPC: E. Adamson[4,5], V. Pizzo[5], D. Biesecker[5],
Model developers: D. Odstrcil[6,1], C. N. Arge[1], C. Henney[7], and S. Wallace[8]

Version 2018/11/10 00:49

ABSTRACT

We report the progress made within the first year of the NASA/NOAA MOU Annex validation project to assess improvements in CME arrival time forecasts using the new ADAPT model, and upgrades to the WSA, and ENLIL models. In year one we focused on the following three tasks: (a) replicating the current operational ENLIL version using the latest ENLIL version, creating a set of benchmark simulations driven by a single map (current operations), (b) performing a new set of simulations using a time-dependent inner boundary from a time-sequence of maps, and (c) a single event (2014-08-15) example ADAPT-WSA-ENLIL simulation. Interim results show that on average, the mean absolute error (MAE) in the arrival time of the time-dependent runs (b) are slightly larger than those of the the benchmark single map runs (a) by $\mathrm{MAE_{TD-bench}}=0.7^{+0.7}_{-0.7}$ hours. Although the error bars show that this difference is not statistically significant there is a trend for a slight increase in error when switching from single map-driven to time-dependent map-driven inputs. We also discuss: challenges in detecting CME arrival in simulations, the comparison of results from zero-point corrected and uncorrected magnetogram inputs for 2007–2017, the comparison of different WSA versions for one single event (2014-08-15), and analysis of the ADAPT-WSA-ENLIL simulation for this event.

Keywords: Validation

1. PROJECT OVERVIEW

In May 2017, National Oceanic and Atmospheric Administration (NOAA) Space Weather Prediction Center (SWPC) and Community Coordinated Modeling Center CCMC started a new project under an annex to a memorandum of understanding between NASA and NOAA.

The purpose of this project is to assess potential improvements in CME arrival time forecasts at Earth afforded by incorporation of time-dependence within the inner boundary conditions of ENLIL and by coupling to the Air Force Data Assimilative Photospheric Flux Transport (ADAPT) model driven by data from the Global Oscillation Network Group *Global Oscillation Network Group* (GONG: Harvey et al. 1996) ground observatories. These outputs are then fed into the coupled Wang-Sheeley-Arge (WSA) - ENLIL model and then compared to the current operational version of WSA-ENLIL (without ADAPT).

The project is performed in close collaboration with model developers Carl Henney, Nick Arge, and Dusan Odstrcil. The community is encouraged to follow the SWPC/CCMC project website which always contains up to date information on the status of the project (ccmc.gsfc.nasa.gov/annex/). All simulations performed in support of this project are available for download from the project website.

[1] NASA Goddard Space Flight Center, Greenbelt, MD, USA
[2] Catholic University of America, Washington, DC, USA
[3] University of Maryland, Baltimore County, USA
[4] Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder, Boulder, CO
[5] NOAA Space Weather Prediction Center, Boulder, CO
[6] George Mason University, Fairfax, Virginia, USA
[7] Air Force Research Laboratory, Kirtland Air Force Base, New Mexico, USA
[8] Department of Physics and Astromony, University of New Mexico

2. METHODOLOGY

Currently, SWPC operational forecasts use WSA version 2.2 and ENLIL version 2.6.2, driven by a single daily-updated zero-point uncorrected GONG map. For the purposes of this project, WSA versions 2.2 and 4.5, and ENLIL version 2.9e have been used.

SWPC has selected a set of 33 historical events (Section 3) to test improvements in CME arrival time prediction. The overall validation project consists of multiple simulation experiments for the entire event set:

a) Benchmark: replicating single GONG map driven WSA version 2.2 and ENLIL version 2.6.2 with ENLIL version 2.9e

b) Time-dependent sequence of GONG maps driving WSA version 2.2 and ENLIL version 2.9e.

c) For a single event, test simulation of: Time-dependent sequence of GONG maps driving ADAPT, WSA version 4.5 and ENLIL version 2.9e.

d) Single GONG map driving ADAPT, WSA version 4.5 and ENLIL version 2.9e.

e) Time-dependent sequence of GONG maps driving ADAPT, WSA version 4.5 and ENLIL version 2.9e.

Items (a)-(c) were selected for year 1 of the project, and are the main focus of this interim year 1 report. To achieve (a), of the 33 events, a subset of 7 events were chosen to fully test that CCMC could replicate the operational ENLIL version 2.6 using ENLIL version 2.9e (Section 4.1).

For each stage, ENLIL settings (Section 4) will be kept constant, and if it is desired to check the effect of updating settings, this will be performed in sub-stages, such that only one setting is changed at a time. After each stage, the performance of the new simulation results will be compared to the benchmark (stage a) and to other previous stages. Specific metrics for this validation study have been selected (Section 5).

WSA version 2.2 will be used for (a) and (b) simulations using regular GONG observations corobs=gongb, but these simulations will later be repeated with WSA version 4.5 for comparison. Note that, compared to version 2.2, WSA version 4.5 uses different coefficients in the velocity equation which are more appropriate for the more recent zero-point corrected GONG observations (corobs=gongz). Therefore, four more variations of (a) and (b) will be performed in year 2:

1. WSA version 2.2, ENLIL version 2.9e with amb=a8b1 settings (newly updated model ambient solar wind; corobs=gongb).
2. WSA version 4.5, ENLIL version 2.9e with amb=a3b2 replication settings; corobs=gongb).
3. WSA version 4.5, ENLIL version 2.9e with amb=a8b1 settings; corobs=gongb).
4. WSA version 4.5, ENLIL version 2.9e with amb=a8b1 replication settings, and zero point corrected magnetogram inputs (corobs=gongz).

In Section 7.1 we test all of these variations for one single CME event case in addition to running ADAPT-WSA-ENLIL simulations for this single event (c). Simulations will be performed using all 12 realizations of ADAPT for stages (c), (d), and (e), and we will provide CME arrival time ensemble analysis of the model output. The ADAPT maps for this study are downloaded directly from ftp://gong2.nso.edu/adapt/maps/gong/ which is the current operational source.

3. EVENTS

SWPC has selected a set of 33 historical events and 36 CME operational input parameters (5 of the events contain two CMEs each) have been provided over the period of three years from 2012–2014. The events are listed in Table 1 with the event number, CME start time, input magnetogram time, CME parameters of latitude, longitude, half-width ($w/2$), and speed (v) in Heliocentric Earth Equatorial (HEEQ) coordinates, and the observed CME arrival time from the *Advanced Composition Explorer* (ACE: Stone et al. (1998)) spacecraft. At least two of the events (2012-05-17 22:00 UT and 2013-05-22 08:48 UT), have "glancing blow"/"flank" arrivals. The other events are generally Earth-directed.

The CME start time is determined from the first time the CME is visible in the *SOlar and Heliospheric Observatory* (SOHO: Domingo et al. (1995)) *Large Angle and Spectrometric Coronagraph Experiment* (LASCO: Brueckner et al. (1995)) or the *Solar TErrestrial RElations Observatory* (STEREO: Kaiser et al. (2008)) *Sun Earth Connection Coronal and Heliospheric Investigation* (SECCHI: Howard et al. 2008) Ahead/Behind coronagraphs. The events and parameters for this study are also available from CCMC's public Space Weather Database Of Notifications, Knowledge, Information (DONKI; ccmc.gsfc.nasa.gov/donki) database via an API and can be downloaded in text format:
https://kauai.ccmc.gsfc.nasa.gov/DONKI/WS/get/CMEAnalysis.txt?startDate=2012-01-01&endDate=2014-12-31&mostAccurateOnly=false&keyword=swpc_annex
or JSON format:
https://kauai.ccmc.gsfc.nasa.gov/DONKI/WS/get/CMEAnalysis?startDate=2012-01-01&endDate=2014-12-31&mostAccurateOnly=false&keyword=swpc_annex

4. MODEL SETTINGS AND INPUTS

4.1. *Replication*

SWPC selected seven different events for which full operational WSA (2.2) ENLIL (2.6.2) model inputs and outputs were available and provided these to CCMC. This event subset is listed in Table 2. We used this event subset to confirm that ENLIL versions 2.9d and 2.9e could replicate the current operational 2.6.2. The replicated runs form the first benchmark for item (a) are described in Section 2. This was done for two reasons: (1) full model outputs were not available for all of the operational runs and (2) this allows us to create variations on the benchmark, such as using WSA 4.5 and corobs=gongz.

The ENLIL model settings for version 2.6.2 important to achieve replication with version 2.9e are listed in Table 3. We tested our replication settings for all 7 events (2) and Figure 1 shows comparisons between the SWPC provided operational model outputs (red) and the ENLIL version 2.9e replicated outputs (black) for MHD quantities at Earth for event #1 (2011-11-22 12:21 UT and 2011-11-26 11:46 UT CMEs). These figures (and the ones for the other events) give us confidence that, for the purposes of CME arrival time, we are able to achieve replication SWPC operational WSA 2.2 and ENLIL 2.6.2. This is further discussed in Section 6.1, where we assess if the benchmark simulations correctly replicate the simulated CME arrival times provided by SWPC.

However, we note that there are some replication nuances that we do not consider in this study, but would be important to achieve exact replication. Eric Adamson has been investigating differences between the different ENLIL versions in how the variables vfast, bfast, dfast, and tfast are used for scaling and/or clipping WSA values. We also note that the general default value of bscl for GONG-driven WSA-ENLIL simulations installed at CCMC, and other installations is bscl=4. Dusan Odstrcil suspects a miscommunication led to using bscl=1 in SWPC operations. Both of the issues, and the time spent on achieving replication for this study, illustrate the importance of code documentation and version control moving forward.

4.2. *Spatial Resolution*

CCMC performed spatial resolution tests for four events (2011-11-22, 2012-01-23, 2012-06-13, 2012-09-28) of the replication event subset listed in Table 2 (marked with the † symbol). For these four events we performed ENLIL simulations at low 256×30×90 (r, θ, ϕ), medium 512×60×180, and high 1024×120×360 resolutions. We

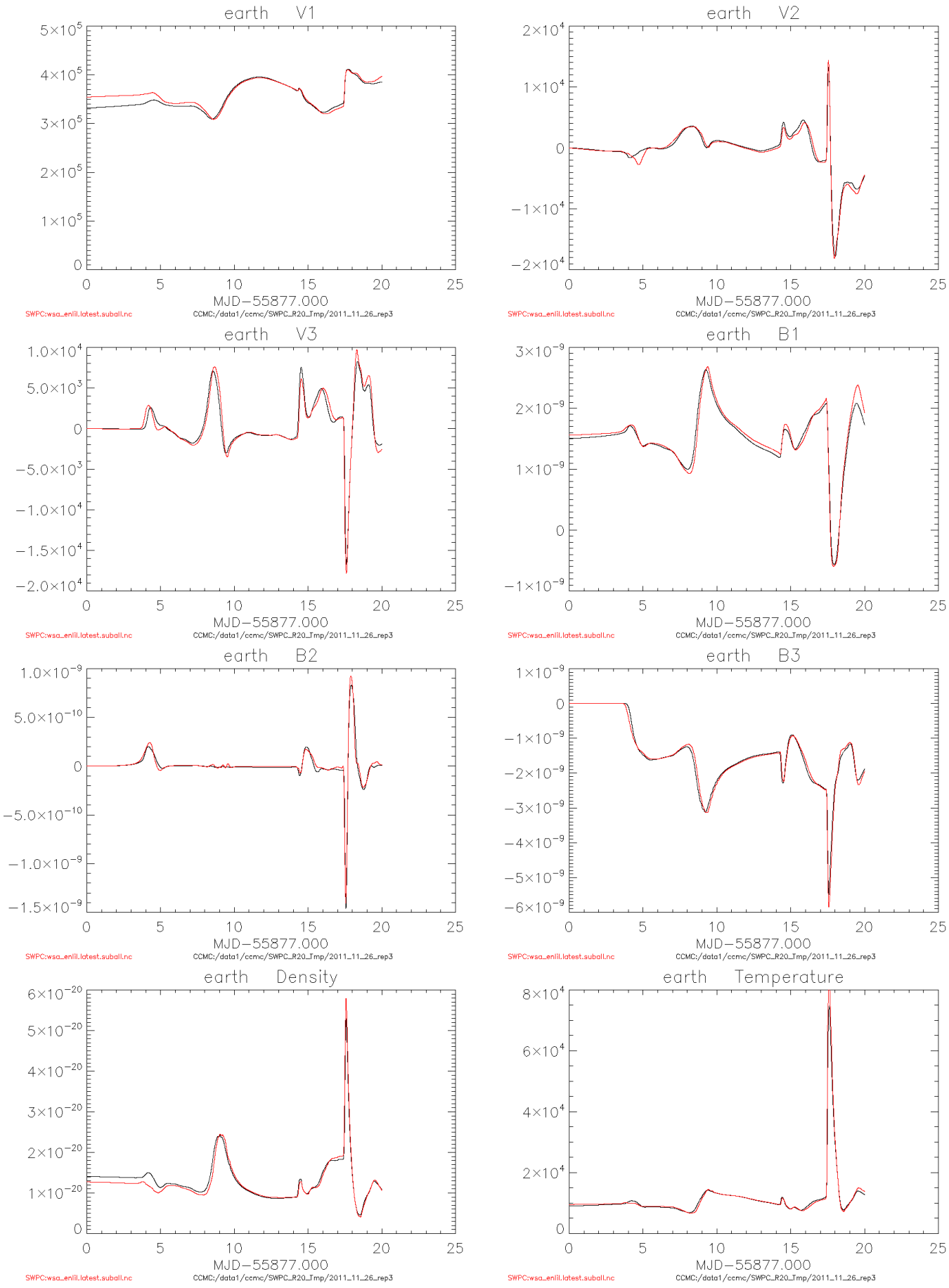

Figure 1. Example of replication results at Earth for Event #1 in Table 2.

Table 1
Summary CME input parameters and observed arrival times for the 33 selected events to be validated, as described in Section 3.

Event #	CME start time [UT]	CME time at 21.5 $R_\odot$ [UT]	Magnetogram Time* [UT]	Lat [°]	Lon [°]	$w/2$ [°]	v [$km\,s^{-1}$]	Arrival Time† [UT]
1	2012-01-23 04:00	2012-01-23 05:50	2012-01-23 06:00	29	17	52	1796	2012-01-24 14:31
2	2012-02-10 20:00	2012-02-11 02:56	2012-02-11 00:00	16	-21	40	518	2012-02-14 07:03
3	2012-02-24 03:46	2012-02-24 08:00	2012-02-24 18:00	23	-8	37	821	2012-02-26 20:58
4	2012-03-07 01:25	2012-03-07 01:59	2012-03-07 06:00	24	-29	50	2040	2012-03-08 10:45
5	2012-04-02 02:12	2012-04-02 10:02	2012-04-02 16:00	26	-14	31	532	2012-04-05 20:03
6	2012-04-18 17:24	2012-04-18 22:25	2012-04-20 04:00	-22	27	36	626	2012-04-21 09:25
	2012-04-19 07:00	2012-04-19 14:55	2012-04-20 04:00	-26	-6	26	313	2012-04-23 02:27
7	2012-05-17 01:48	2012-05-17 04:34	2012-05-17 22:00	-17	71	54	1263	2012-05-20 01:36
8	2012-06-13 13:25	2012-06-13 20:00	2012-06-14 18:00	-36	-33	49	597	2012-06-16 09:01
	2012-06-14 14:12	2012-06-14 16:57	2012-06-14 18:00	-27	-6	47	1177	2012-06-16 19:31
9	2012-07-12 16:36	2012-07-12 19:13	2012-07-13 04:00	-14	-1	55	1453	2012-07-14 17:28
10	2012-07-28 21:12	2012-07-29 05:25	2012-07-29 04:00	-21	-37	43	382	2012-08-02 09:22
11	2012-08-31 20:00	2012-08-31 22:46	2012-09-01 06:00	6	-30	33	1010	2012-09-03 11:23
12	2012-09-28 00:00	2012-09-28 03:49	2012-09-28 18:00	10	20	55	872	2012-09-30 22:13
13	2012-10-05 02:48	2012-10-05 08:47	2012-10-05 18:00	-18	7	42	698	2012-10-08 04:31
14	2012-10-27 12:48	2012-10-28 00:28	2012-10-28 22:00	6	11	30	375	2012-10-31 14:40
15	2012-11-20 12:00	2012-11-20 17:40	2012-11-21 00:00	19	24	47	664	2012-11-23 21:12
16	2013-01-13 06:00	2013-01-13 19:02	2013-01-13 20:00	1	-3	36	463	2013-01-16 23:00
17	2013-01-13 06:00	2013-01-13 19:02	2013-01-15 20:00	1	-3	36	463	2013-01-16 23:00
	2013-01-15 08:48	2013-01-15 16:26	2013-01-15 20:00	-27	15	23	428	2013-01-19 16:47
18	2013-03-12 10:36	2013-03-12 15:24	2013-03-12 18:00	42	8	52	734	2013-03-15 05:05
19	2013-03-15 07:12	2013-03-15 09:34	2013-03-15 10:00	-9	-5	55	1399	2013-03-17 05:28
20	2013-04-11 07:24	2013-04-11 11:54	2013-04-11 12:00	-6	-17	48	743	2013-04-13 22:15
21	2013-04-26 18:24	2013-04-27 04:37	2013-04-29 16:00	-32	8	30	493	2013-04-30 08:57
22	2013-05-17 09:12	2013-05-17 11:22	2013-05-17 18:00	6	-24	40	1498	2013-05-19 22:21
23	2013-05-22 08:48	2013-05-22 15:37	2013-05-22 20:00	9	49	55	1488	2013-05-24 17:35
24	2013-06-28 02:00	2013-06-28 05:09	2013-06-28 16:00	-29	42	45	1063	2013-06-30 10:40
25	2013-07-06 19:36	2013-07-06 22:05	2013-07-07 14:00	2	4	42	560	2013-07-09 19:58
26	2014-01-04 19:54	2014-01-04 23:12	2014-01-06 18:00	-38	6	42	806	2014-01-07 14:25
27	2014-01-07 18:24	2014-01-07 20:01	2014-01-08 00:00	-21	21	50	2048	2014-01-09 19:31
28	2014-03-23 03:36	2014-03-23 08:34	2014-03-23 18:00	3	-53	47	768	2014-03-25 19:25
29	2014-04-17 01:25	2014-04-17 07:07	2014-04-18 18:00	-26	15	38	529	2014-04-20 10:24
30‡	2014-08-15 17:48	2014-08-16 02:14	2014-08-16 16:00	12	2	38	438	2014-08-19 05:58
	2014-04-17 12:48	2014-04-18 16:14	2014-04-18 18:00	-18	17	43	1043	2014-04-20 10:24
31	2014-09-02 16:48	2014-09-02 22:17	2014-09-03 14:00	37	-9	37	708	2014-09-06 14:19
32	2014-09-09 00:06	2014-09-09 04:05	2014-09-09 16:00	24	-23	43	767	2014-09-11 22:58
33	2014-09-09 00:06	2014-09-09 04:05	2014-09-11 02:00	24	-23	43	767	2014-09-11 22:58
	2014-09-10 18:00	2014-09-10 20:16	2014-09-11 02:00	15	2	45	1343	2014-09-12 15:30

* Observation time of latest magnetogram. Applicable to single map-driven runs only.
† CME-associated shock observed by the ACE spacecraft.
‡ CME event chosen for example ADAPT-WSA-ENLIL simulation with 12 realizations of ADAPT (item c).

Table 2
Seven events selected by SWPC to test ENLIL version 2.6.2 replication.

Event	CME time at 21.5 $R_\odot$ [UT]	Lat [°]	Lon [°]	$w/2$ [°]	v [km s^{-1}]
1†	2011-11-22 12:21	12	-46	34	559
	2011-11-26 11:46	8	52	55	790
2†	2012-01-23 05:50	29	17	52	1796
3	2012-03-13 19:48	18	55	51	1512
4†	2012-06-13 20:00	-36	-33	49	597
	2012-06-14 16:57	-27	-6	47	1177
5	2012-08-31 22:57	5	-27	35	1002
6†	2012-09-28 03:49	10	20	55	872
7	2012-11-20 17:40	19	24	47	664

† Resolution tests were also performed for these events.

additionally simulated the 2011-11-22 event at high×2 resolution 2048×240×720. Figure 2 shows the model output time-series of speed, density, magnetic field, and temperature at Earth for different spatial resolutions for each event. SWPC inspected these resolution tests and chose to continue at medium resolution for the validation study of the 33 events (listed in Table 1), based on the sharpness in the CME arrival compared to low resolution, and the small difference in CME arrival compared to high resolution.

4.3. *Time resolution*

SWPC decided to use a 1 hour time resolution for the 3D output time step, and 1-3 minute output at locations of interest. This is the output setting for the current operational ENLIL version 2.6.2 at SWPC.

4.4. *Magnetogram Inputs to WSA*

For the benchmark runs, the simulations will be driven by a single zero-point uncorrected GONG map provided by SWPC. The WSA velocity output file from each map that was used operationally to model the CME event in real-time was provided to us, and was used as input for the benchmark simulations. For time-dependent runs, it is necessary to choose the time-interval for updating the magnetogram at the model inner boundary. Hourly daily-updated magnetograms are available after November 2012, and prior to this date they are available every 6 hours. For settings to be uniform for all runs, SWPC has chosen a magnetogram update frequency of 6 hours for the validation study. However, we will perform a series of test runs to test the effects of magnetogram time cadence on a subset of events. All of the runs for (a) and (b) in this report use a uniform 6 hour input time-cadence, and (c) uses a 2 hour time-cadence for ADAPT.

Model developer Dusan Odstrcil has provided a comparison of the background solar wind prediction for the 2007-2017 time period when using zero-point corrected maps (gongz) or uncorrected maps (gongb), both with a 24 hour input cadence. Figures S1-S2 shows the differences in the model output for these two input options. Overall, the magnetic field magnitude is better captured when using the zero-point corrected maps. The four variations of tasks (a) and (b) to be performed in year 2 (Section 2) include performing the same simulations using zero-point corrected maps and this is further discussed in Section 7.1.

5. METRICS

Starting in September 2017, SWPC and CCMC began discussing metrics. SWPC determined that the main metrics for this validation study will be the Mean Error (ME), Mean Absolute Error (MAE), and the Root Mean Square Error (RMSE). We define the CME arrival time prediction error Δt for each particular forecast as

$$\Delta t = t^{fcst} - t^{obs}, \quad (1)$$

where t^{fcst} is the predicted CME arrival time and t^{obs} is the observed CME arrival time and is computed for hit events. This follows the standard practice from atmospheric sciences (Jolliffe & Stephenson 2011) and the CME Arrival Time Scoreboard (kauai.ccmc.gsfc.nasa.gov/CMEscoreboard/). A negative Δt corresponds to a CME arrival observed later than predicted and a positive Δt, to an arrival observed earlier than predicted.

In the context of contingency tables, a hit is defined when a CME is both predicted and observed to arrive. A false alarm is when a CME is predicted to arrive but is not observed to arrive. A miss is when a CME is observed to arrive but it was not predicted to arrive. Finally, a correct rejection (or correct negative) is when a CME is both not predicted and not observed to arrive. Because this event set begins with hit events in the benchmark, generally only the metrics on hits are valid for this study (not contingency table skill scores). However, we will keep track of hits in the benchmark set that become misses in subsequent simulations.

The observed arrival times t^{obs} at ACE used for this validation study were provided by SWPC and are listed in the last column of Table 1 and are also available in the DONKI database. Note that we are comparing model outputs at the Earth trajectory location to ACE observations at L1, about 0.1 AU from the Earth along the Sun-Earth line. With the medium resolution grid, it takes about 0.08–0.2 hours for CMEs 400–1000 km s^{-1} to traverse one grid cell in the radial direction, and about 0.4–1 hours to traverse the grid cells between L1 and the Earth. Therefore we do not expect the CME arrival time error Δt from these runs to be less than 1 hour on average.

SWPC and CCMC have agreed to use an existing CCMC algorithm (briefly described in Wold et al. (2018)) to derive the arrival time t^{fcst} from the simulation time series that checks for when the time derivative of the simulated dynamic pressure exceeds 3 nPa. This detects the beginning of the rise in simulated speed and density related to the CME arrival. SWPC generally determines the arrival time from the simulation from the mid-point of the rise visible in the speed. For the purposes of this project these two methods are not expected to produce very different results compared to the overall arrival time errors. Note that operationally, SWPC forecasters adjust the predicted arrival time t^{fcst} based on uncertainties before issuing a forecast, and we do not consider these human-in-the-loop forecasts here.

In general, the dynamic pressure threshold algorithm detects the simulated CME arrival as expected. However, in the process of performing this study, we found that the cases where the only the "flank" of the CME

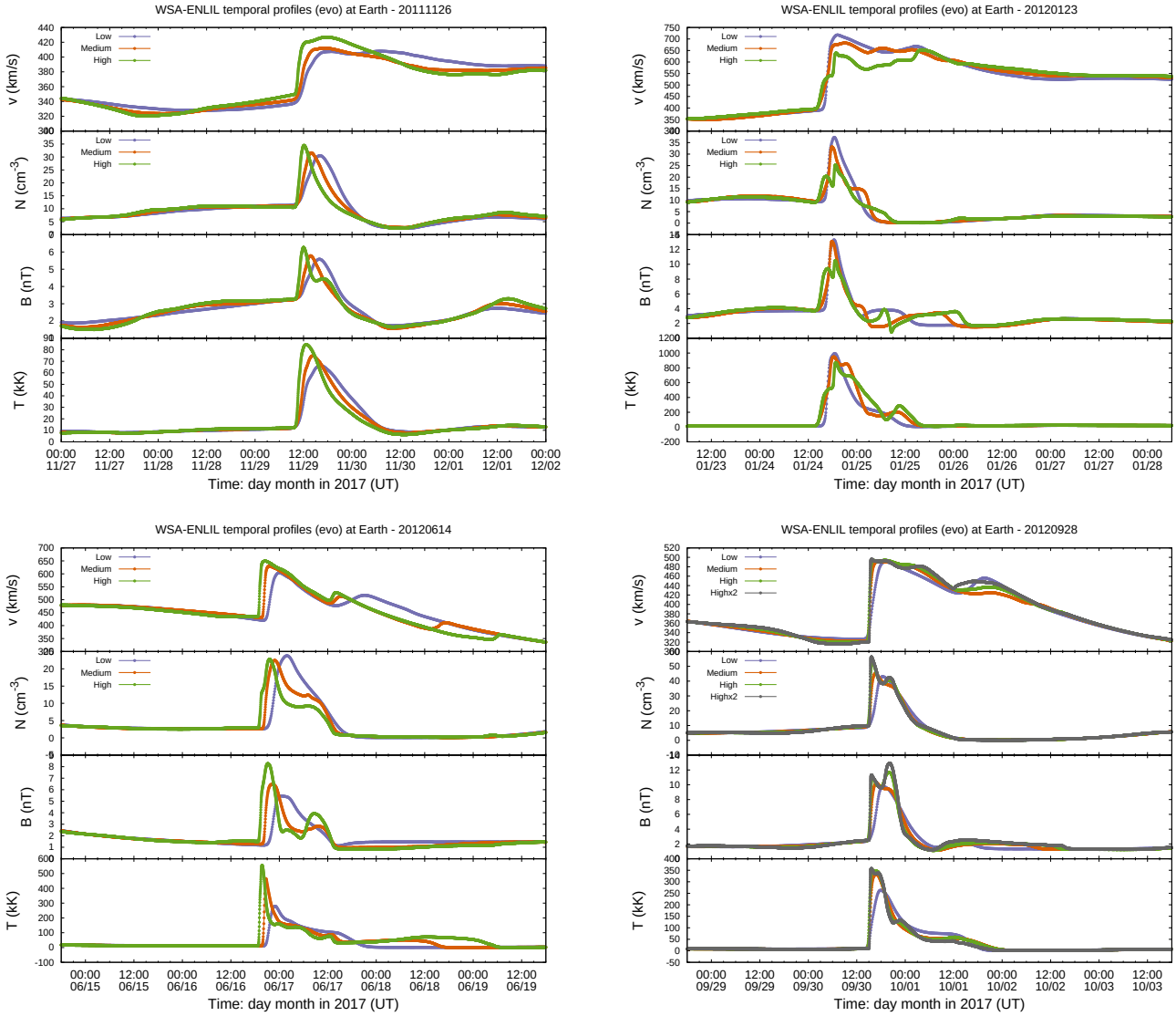

Figure 2. Model output time-series of speed, density, magnetic field, and temperature for different resolution settings for the 4 events (of the 7 events listed in Table 2). Low resolution–purple, medium resolution–orange, high resolution–green, high×2 resolution–grey.

is simulated to arrive at Earth, the algorithm does not detect the CME arrival. Additionally, for one event (2012-04-02) the CME is simulated to have a head-on arrival but has a weak arrival signature that was not detected. And finally, of the five events that contain 2 CMEs each, the 2nd CME arrival was not detected in three of them. In the interests of keeping events that are not only head-on arrivals, we kept these "flank" events in the validation study and instead experimented with decreasing the threshold in the algorithm to 0.5 nPa/hr and adding extra filter conditions. However, even with the adjustments, the algorithm fails to detect arrivals for two events (2012-04-18 and 2012-06-13). The final pressure algorithm with filters, developed by Jan Merka, consists of:

1. Search for CMEs by looking for increasing values of dp (cloud tracer variable). Keep track of the maximum dp for the CME and search for the next CME.
2. For each detected CME, search for the CME arrival for 6 hours before the dp rise up until the maximum dp for that CME.
3. The CME arrival time is detected when the time derivative of the dynamic pressure $P_{\rm dyn} = m_p n v^2$ exceeds the threshold of 0.5 nPa/hr, and the temperature and magnetic field derivatives are zero or positive.
4. Search for the next CME arrival time only after the derivative of the dynamic pressure is negative (below -0.0075 nPa/hr).
5. If no CME arrival is detected for CME(s) detected by dp, repeat the arrival time search once more, using a lower threshold of 0.1 nPa.

However, while these algorithm adjustments worked somewhat well for studies (a) and (b), we wanted a robust, automated method that will work for the entire project. Especially when there will be 12 arrivals to detect for each CME when using ADAPT inputs. For

Table 3
ENLIL model settings for version 2.6.2 important for replication with version 2.9e.

Variable	Value	Description
amb	**a3b2**	**Ambient wind conditions setting:**
nbrad	3	Magnetic field correction
bfast	300	Radial magnetic field of fast stream (nT)
bslow	0	Radial magnetic field of slow stream (nT)
bmean	0	Radial magnetic field of mean stream (nT)
bscl	1	Magnetic field scaling factor
dfast	200	Number density of fast stream (cm^{-3})
dslow	2000	Number density of slow stream (cm^{-3})
dmean	0	Number density of mean stream (cm^{-3})
dscl	1	Number density scaling factor
tfast	0.08	Mean temperature of fast stream (MK)
tslow	0.1	Mean temperature of slow stream (MK)
tmean	0	Mean temperature of mean stream (MK)
tscl	1	Mean temperature scaling factor
vfast	675	Radial flow velocity of fast stream ($km\,s^{-1}$)
vslow	225	Radial flow velocity of slow stream ($km\,s^{-1}$)
vmean	0	Radial flow velocity of mean stream ($km\,s^{-1}$)
vrfast	25	Reduction of the maximum flow velocity ($km\,s^{-1}$)
vrslow	25	Reduction of the minimum flow velocity ($km\,s^{-1}$)
shift	8	Azimuthal shift at the inner boundary (deg)
nshift	1	Azimuthal shift at the inner boundary
xalpha	0	Fraction of alpha particles (rel to protons)
dvexp	2	Exponent in NV^{dvexp} constant condition
nptot	0	0 if P_{the} balance at boundary
Numerical model setting		
difbb	nodifbb	Magnetic field diffusion
cdifbb	1	Div(B) diffusion coefficient
cdifb	0.4	Div(B) diffusion coefficient
ceclip	0.01	Clip the thermal energy below fraction of the total energy
cvetot	-1	Compressive velocity threshold for Etot
vc	vc_avr	Characteristic speed at cell interfaces
vs	vs_max	Signal speed at cell interfaces
bchalf	nobchalf	Boundary conditions at the half step
divb	nodivb	Div(B) treament
mftrace	mfeuler	IMF line tracing
gamma and heating: heat	noheat	Global heating
gamma	1.6666667	Ratio of specific heats
qheat	0	Volumetric heating factor
nheat	0	Volumetric heating type
Spatial grid and timing		
x1l	0.1	Inner boundary (AU)
x1r	1.7	Outer boundary (AU)
tfrom	-48.	Output from this time (h)
tstart	-360	Start computation at this time (h)
tstop	120	Stop computation at this time (h)
tstep	1	Output with this step in time (h)
res	med	Numerical grid resolution
cormode	single	Coronal data mode
corobs	gongb	Observatory name

this reason, we followed ENLIL model developer Dusan Odstrcil's recommendation of running the simulations on 2 "blocks" (nblk=2), wherein one simulation block contains the ambient simulation, and the other contains the ambient + CME(s) simulation. The initial algorithm from Odstrcil checks for a 25% increase between the two simulation blocks in the density time-series, and detects the arrival time as the inflection point after this criteria is reached. This algorithm proved to be more robust than the pressure-threshold algorithm (using one simulation block), however, there are still issues detecting the 2nd CME in a few of the events (2012-04-18, 2014-04-17, 2014-09-09). We worked with Odstrcil to slightly adjust the nblk=2 algorithm to the following:

1. Search for the timestamp when the ratio of CME simulation density to the ambient simulation density is greater than 1.10: $n_{\rm CME}/n_{\rm ambient} > 1.10$.

2. Next, the inflection point is found immediately following the timestamp from step 1. This is recorded as the CME arrival time.

3. Find the maximum difference between CME and ambient simulation densities after arrival time detection. Do not search for another CME arrival until difference between the CME and ambient simulation densities start to increase again after this: $n_{\rm CME}(t)/n_{\rm ambient}(t) > n_{\rm CME}(t-1)/n_{\rm ambient}(t-1)$

4. Before resuming the search after the timestamp found in step 3, first check that CME simulation density goes back down to the ambient simulation density ($n_{\rm CME}/n_{\rm ambient} \leq 0$).

5. Continue searching for the next CME arrival, starting at step 1.

Finally, we settled on a simpler multi-block algorithm, by running all of the simulations on nblk= $n_{\rm CMEs} + 1$ blocks. For our study, this means all of the single CME simulations were performed with nblk=2, and all of the double CME simulations were performed with nblk=3. For nblk=3: block 2 is for the ambient simulation (no CMEs), block 1, the simulation for the first CME, and block 0, the simulation for the first and second CMEs together (all CMEs). Still, one 2 CME simulation (2014-04-17) had such a small density signature for the 2nd CME arrival, that we changed the algorithm to use the simulation temperature instead of density for all events. The drawback of both the nblk=2 and nblk=3 methods is the extra computational resources and storage required. The multi-block algorithm is as follows:

For the 1st CME:

1. Search for the timestamp when the ratio of CME 1+2 simulation temperature (block 0) to the ambient simulation temperature is greater than 1.10: $T_{\rm CME1+2}/T_{\rm ambient} > 1.10$.

2. Next, the inflection point is found immediately following the timestamp from step 1. This is recorded as the 1st CME arrival time.

For the 2nd CME (if nblk=3):

4. Search for the timestamp when the ratio of CME 2 simulation temperature to the ambient simulation temperature is greater than 1.10: $(T_{\rm CME1+2} - T_{\rm CME1} + T_{\rm ambient})/T_{\rm ambient} > 1.10$.

5. Next, the inflection point is found immediately following the timestamp from step 3. This is recorded as the 2nd CME arrival time.

Our work has shown that it is non-trivial to reliably, automatically detect CME arrivals in the model output time-series using any of these algorithms (pressure threshold, nblk=2 density, multi-block temperature). If necessary, another algorithm may be developed, perhaps taking advantage of the 3D model output. For the time being, we have chosen to use the multi-block temperature algorithm for the validation performed in this report, which we have demonstrated will automatically work for all of the events. If a new method is developed, we will rerun the CME arrival detection so that all of the simulations in the project will be treated uniformly.

Now that t^{obs} and t^{fcst} have been defined for computing Δt, next we outline the metrics. The first metric is the Mean Error (ME), given by

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

where N is the total number of hit events in the validation set consisting of each event i. The Mean Error is a way of quantifying the bias, as it describes the models tendency to consistently predict early or late arrival with respect to observations. A negative bias corresponds to on average early predicted arrivals, while a positive bias corresponds to on average late predicted arrivals.

The Mean Absolute Error (MAE) is defined as

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

While the ME is a measure of the model's bias, it is not adequate to measure the forecasting skill of a model, since negative errors can compensate positive errors. By taking the absolute value of the errors and measuring the distance between the observed and forecast values, we can overcome this. The MAE is very similar to the Mean Square Error (MSE) and the Root Mean Square Error (RMSE), but does have some differences. In practice, it is also more resistant to outlier errors.

The Root Mean Square Error (RMSE) is a second order moment, and is given by

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

Compared to the MAE, the RMSE gives more weight to larger errors, due to the errors being squared.

SWPC and CCMC also discussed the measuring forecast performance for other quantities such as comparing the ICME sheath observed mean or max plasma quantities to the simulated quantities, or the magnetic cloud duration. More discussion is needed this regard on how to make such comparisons and what to compare. Currently, this is outside the scope of this project, however,

all of the simulations from this study will be available and we expect these comparisons will be made in follow-up studies.

6. YEAR 1 RESULTS

6.1. *Benchmark Simulations (a)*

To complete item (a), a set of benchmark simulations were performed for the 33 events in this study. To create the benchmark, we used the operational WSA velocity files and the CME parameters listed in Table 1) as input and replicated ENLIL version 2.6.2 using version 2.9e, as discussed in Section 4.1. Next, we compared the SWPC-provided manually identified simulated arrival time from the operational runs with the multi-block temperature algorithm identified simulated arrival time from the benchmark runs (see Section 5). Comparing these two simulated arrival times (algorithm identified arrival time−manually identified arrival time) yields a mean error of $\mathrm{ME}_{\mathrm{algorithm-manual}}$=+0.4 hours and mean absolute error of $\mathrm{MAE}_{\mathrm{algorithm-manual}}$=1.3 hours. These errors were determined to be reasonably small to assume that replication was successful, and that the difference in general is due to the algorithm detecting first inflection point of the rise due to the CME arrival, compared to the SWPC manual method of using the mid-point of the rise. There are two outliers in which the benchmark time is very different (2012-04-02 and 2012-04-19) but these are likely due to the manual detection method choosing a different aspect of the time series compared to the algorithm that has the benefit of the background simulation. A full list of the simulation arrival times and prediction errors for each event are listed in Table S1.

In Figures S3-S5 the SWPC provided simulated arrival times (magenta vertical lines) and those detected by the algorithm (purple vertical lines) are shown for comparison on the same plot for each of the 33 events. The benchmark simulation and arrival times are plotted in purple and the one hour average OMNI data is plotted in black.

We computed basic validation metrics (as discussed in Section 5 for the benchmark simulations compared to the ACE observed arrival times provided by SWPC. We found a mean error of ME=-0.2 hours, a mean absolute error of MAE=5.1 hours, and root mean-squared error of RMSE=6.0 hours. These values are shown in Table 4 as (a) Benchmark, along with errors when using the SWPC provided manually identified arrival time from the original set of operational simulations.

6.2. *Time-dependent Simulations (b)*

The same benchmark model settings (Table 3) were applied to the time-dependent runs, except for the setting of cormode=multi up until the simulation rundate (CME time at 21.5 $R_\odot$, after relaxation). With this setting, a time-sequence of WSA velocity output maps from a time-sequence of GONG magnetograms are used to create the simulation inner boundary. Interpolation between velocity maps is performed within ENLIL at each numerical timsestep using a "co-rotating weighted linear interpolation" method for the computational cells at the inner boundary (θ, ϕ). For each numerical timestep, the co-rotating speed in ϕ of the immediately preceding and following maps is used to determine (θ, ϕ) grid cell location on each map, and an area-weighting procedure is used to calculate the interpolated values.

Figures S3-S5 show the simulation time-series for the benchmark single map-driven runs (a) and time-dependent map-driven runs (b) for all 33 events. The simulated benchmark time-series are plotted in purple, time-dependent runs in brown, and the one hour average OMNI data is plotted in black. The magenta vertical lines show the SWPC provided simulated arrival times, the brown vertical lines show those detected by the multi-block algorithm for the time-dependent runs, and the black vertical lines show the ACE observed arrival times (provided by SWPC). Overall, the effects of time-dependence show more variation in the background solar wind compared to the benchmark single-map driven background solar wind.

Similar to the benchmark runs, we also compute validation metrics for the time-dependent runs compared to the ACE observed arrival times provided by SWPC. We found a mean error of ME=+0.6 hours, a mean absolute error of MAE=5.8 hours, and root mean-squared error of RMSE=6.9 hours (see Table 4).

Table 4
CME arrival time error validation results. The MAE of the time-dependent runs are slightly larger than those of the the benchmark single map runs by $\mathrm{MAE}_{\mathrm{TD-bench}}=0.7\ \mathrm{hours}^{+0.7}_{-0.7}$.

Simulation	ME (hrs)	MAE (hrs)	RMSE (hrs)
SWPC operational†	-0.7	4.5	5.6
(a) Benchmark	-0.2±1.9	5.1±1.0	$6.0^{+0.9}_{-1.0}$
(b) Time-dependent	$+0.6^{+2.2}_{-2.1}$	5.8±1.2	$6.9^{+1.2}_{-1.1}$

† Metrics computed using the SWPC provided simulated arrival times from their operational runs.

6.3. *Validation Discussion*

The validation results of the benchmark and time-dependent simulations are summarized in Table 4. We have also visualized the arrival time errors separately for each event and simulation type in Figure 3. In the supplemental plots at the end of the report we also show the predicted vs observed transit time in Figure S6 and the arrival time error vs CME input parameters in Figure S7. We have added 95% confidence intervals to these errors using a bootstrap technique with 10,000 samples using replacement. On average, the MAE of the time-dependent runs are slightly larger than those of the the benchmark single map runs by $\mathrm{MAE}_{\mathrm{TD-bench}}=0.7$ hours. On the other hand, the bias slightly increases from -0.2 hours to +0.6 hours. When we compare the difference in the error to the size of the error bars $\mathrm{MAE}_{\mathrm{TD-bench}}=0.7^{+0.7}_{-0.7}$, the difference is not statistically significant. However, there is a trend for a slight increase in error (0.7 hours) when switching from single map-driven to time-dependent map-driven inputs.

7. OUTLOOK FOR YEAR 2

7.1. *Test ADAPT-driven Simulation (c)*

The final item (c) for year 1 of our project is a test run for a single event using a time-dependent sequence

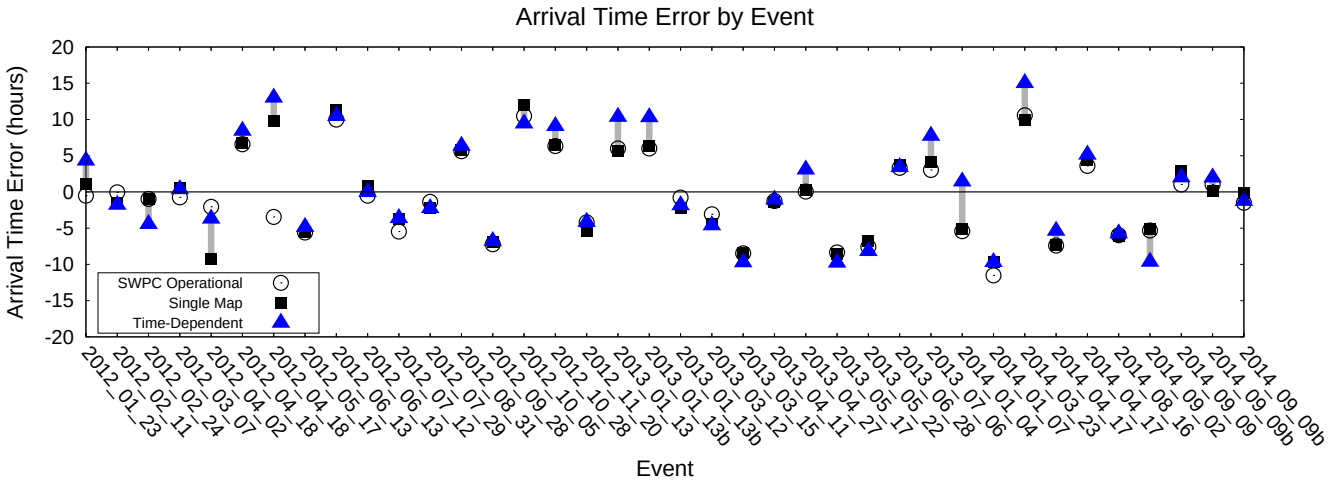

Figure 3. Arrival time error for each event and simulation setting. Circles=SWPC operational arrival times, Square=benchmark single-map driven arrival times, Triangle=time-dependent arrival times. The grey bars show the distance between the single map and time-dependent map driven arrival times.

of GONG maps driving ADAPT, WSA version 4.5 and ENLIL version 2.9e with amb=a8b1 settings. For this test run SWPC has chosen the 2014-08-15 17:48 UT CME event. Figure 4 shows a WSA-ENLIL radial velocity contour plot of the CME for the (a) constant Earth HEEQ latitude plane, (b) meridional plane of Earth, and (c) 1 AU sphere in cylindrical projection on 2014-08-19 04:00 UT. The figure shows that this slow (438 km s^{-1}) CME becomes embedded in slow background solar wind that is faster than the CME itself.

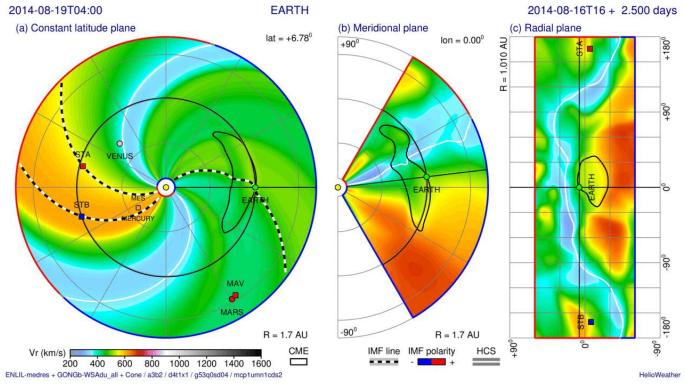

Figure 4. WSA-ENLIL radial velocity contour plot of the CME for the (a) constant Earth HEEQ latitude plane, (b) meridional plane of Earth, and (c) 1 AU sphere in cylindrical projection on 2014-08-19 04:00 UT.

Before embarking on simulating a CME event we worked with model developers Nick Arge, Carl Henney, and Dusan Odstrcil to perform a one month long test run of the ambient solar wind for July–August 2010. This process allowed us to work out any code input/output issues between the models, and test model settings. Figure S8 provided by Dusan Odstrcil shows the 12 resulting ENLIL radial velocity model outputs at different radial distances (0.1, 0.25, 0.5, 1 AU) for July 2010 using a 24 hour ADAPT-WSA input cadence (left) or 2 hour input cadence (right). Our main lesson from this exercise was that running ENLIL at low resolution had the effects of averaging out the variation provided by the 12 ADAPT realizations due to interpolation at the inner boundary. Moving forward, ENLIL runs using ADAPT should be performed at medium resolution or higher. Secondly, we found that using a higher input time-cadence of the ADAPT-WSA at the ENLIL inner boundary can have the effect of smoothing out differences in the realizations. We will investigate this further in year 2 of the project.

Next, we compared WSA version 2.2 and 4.5 with regular GONG daily-update maps (gongb) and zero-point corrected maps (gongz) using cormode=multi for the 2014-08-15 CME. In Figure 5 we show results for ENLIL version 2.9e with amb=a8b1 settings using WSA version 2.2 with gongb (light green) and gongz (dark green) inputs, and similarly when using WSA version 4.5 with gongb (light blue) and gongz (dark blue) inputs. Based on the simulations for this single CME event, both gongb and gongz inputs give similar CME arrival time results from WSA versions 2.2 and 4.5. As discussed in Section 2 we will perform these types of model version and input variations for all of the events in the study.

Finally, we performed 11 ADAPT-WSA-ENLIL simulations (one for each realization) for the 2014-08-15 17:48 UT CME event, using a 2 hour ADAPT-WSA cadence at the ENLIL inner boundary. The ENLIL time-series at Earth for each realization are plotted in shades of green in Figure 6, and the dashed lines show the ambient simulation (with no CME). Note that we do not show results for realization R001 as there were problems with the ADAPT-WSA files for this realization that Nick Arge hopes to resolve soon. Due to the slow speed of this CME, there is not a strong arrival signature in the time-series. The spread in CME arrival times ranges from 2014-08-19 11:21 to 13:11 UT (1.9 hours) with an average arrival time of 2014-08-19 12:20 UT. This gives an average arrival time error of Δt=6.4$^{+0.9}_{-1.0}$ hours, about 12 hours later than the benchmark and time-dependent runs $\Delta t_{\rm benchmark}$=-6.1 hours, $\Delta t_{\rm TD}$=-5.8 hours. This is due to the different ENLIL model ambient settings, which we explore next.

We also investigated the difference in arrival times arising from grid resolution and different ambient solar wind settings and resolutions for one realization, R000.

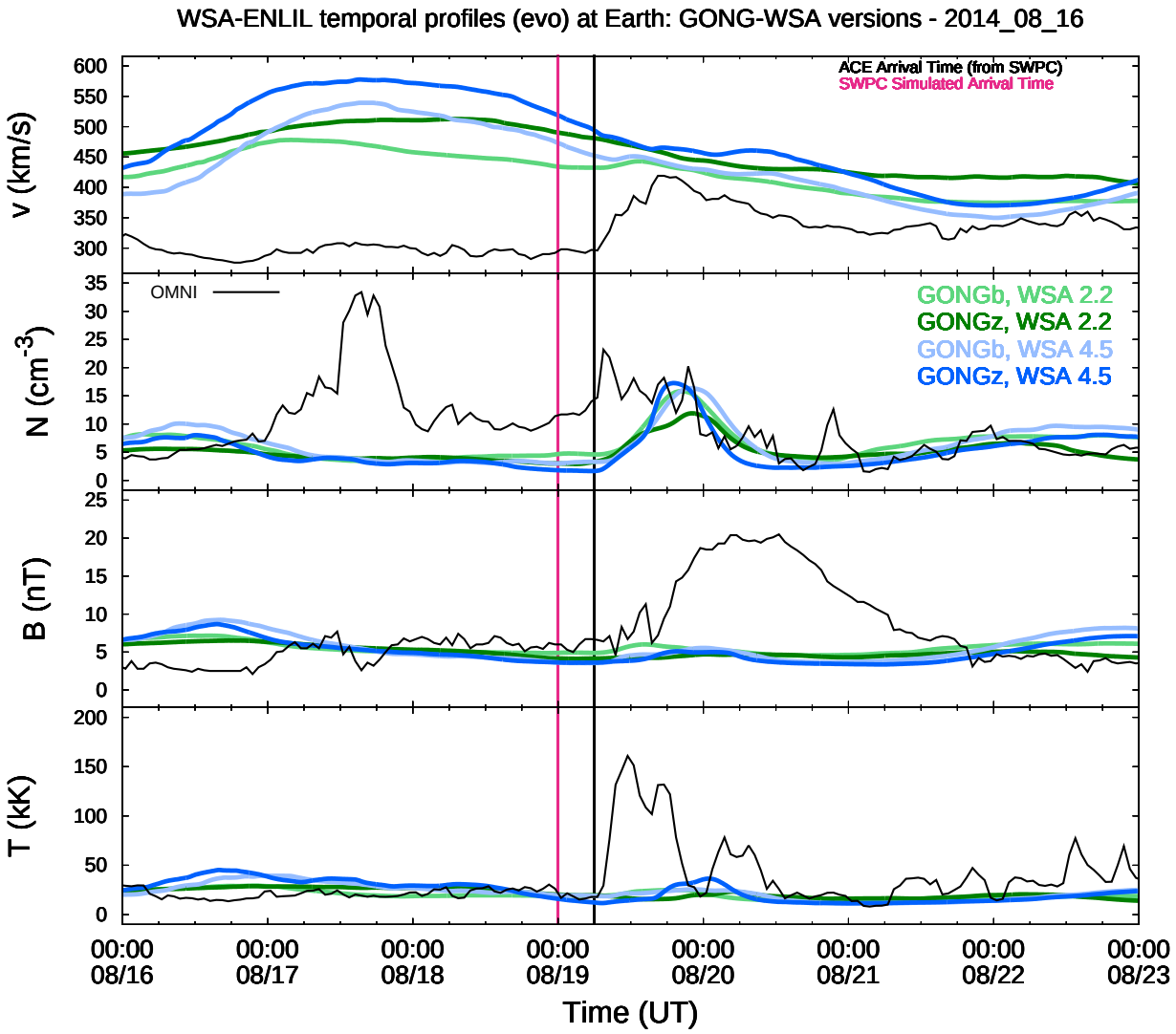

Figure 5. ENLIL version 2.9e results using WSA version 2.2 with gongb (light green) and gongz (dark green) inputs and WSA version 4.5 with gongb (light blue) and gongz (dark blue) inputs. Based on the simulations for this single CME event, gongb and gongz inputs give similar CME arrival time results from WSA versions 2.2 and 4.5.

Figure 7 (left) shows the time-series for low (purple), medium (orange) and high (green) resolution grids for this CME event for R000 (amb=a8b1). These results show a 7.3 hour difference in arrival time between the low and high resolution runs. Such a large difference is likely due to the slow initial CME speed, and the effect of the resolution on smoothing the time-varying background from ADAPT-WSA. Figure 7 (right) shows the simulation time-series for different ambient settings: amb=a8b1 (yellow), amb=a3b2 (blue), and amb=a6b1 (green). There is a 10 hour CME arrival time difference range between the a3b2 to a8b1 ambient settings. As described in Section 2 we will test these ambient settings for all of the events in year 2.

To illustrate the difference in the ENLIL inner boundary conditions when using ADAPT, we show equatorial slices of the ENLIL inner boundary as a function of time in panel (b) of Figure S9. The top figure shows the single map-driven case (cormode=single), the middle figure shows the time-dependent map-driven case (cormode=multi), and the bottom figure shows the time-dependent ADAPT map-driven case for realization R000 (cormode=multi). This figure illustrates that for the single map-driven case (bottom), the background map is fixed and rotated through time during CME insertion, whereas for the time-dependent sequence of maps (middle, bottom), the CME is inserted into a clearly time-varying background.

7.2. *Community Involvement*

Methods and lessons learned from the CME Arrival Time and Impact Team (Verbeke et al. 2018) will be applied to this SWPC/CCMC project, and vice versa. We are also collaborating with the UK MetOffice who plan to perform ADAPT simulations for the same events and input parameters. Additionally, the SWPC/CCMC project's set of operational parameters for 33 events will also be used as a validation test set by team members from the research community. This opens up the possibility for other flux transport, coronal, and heliospheric

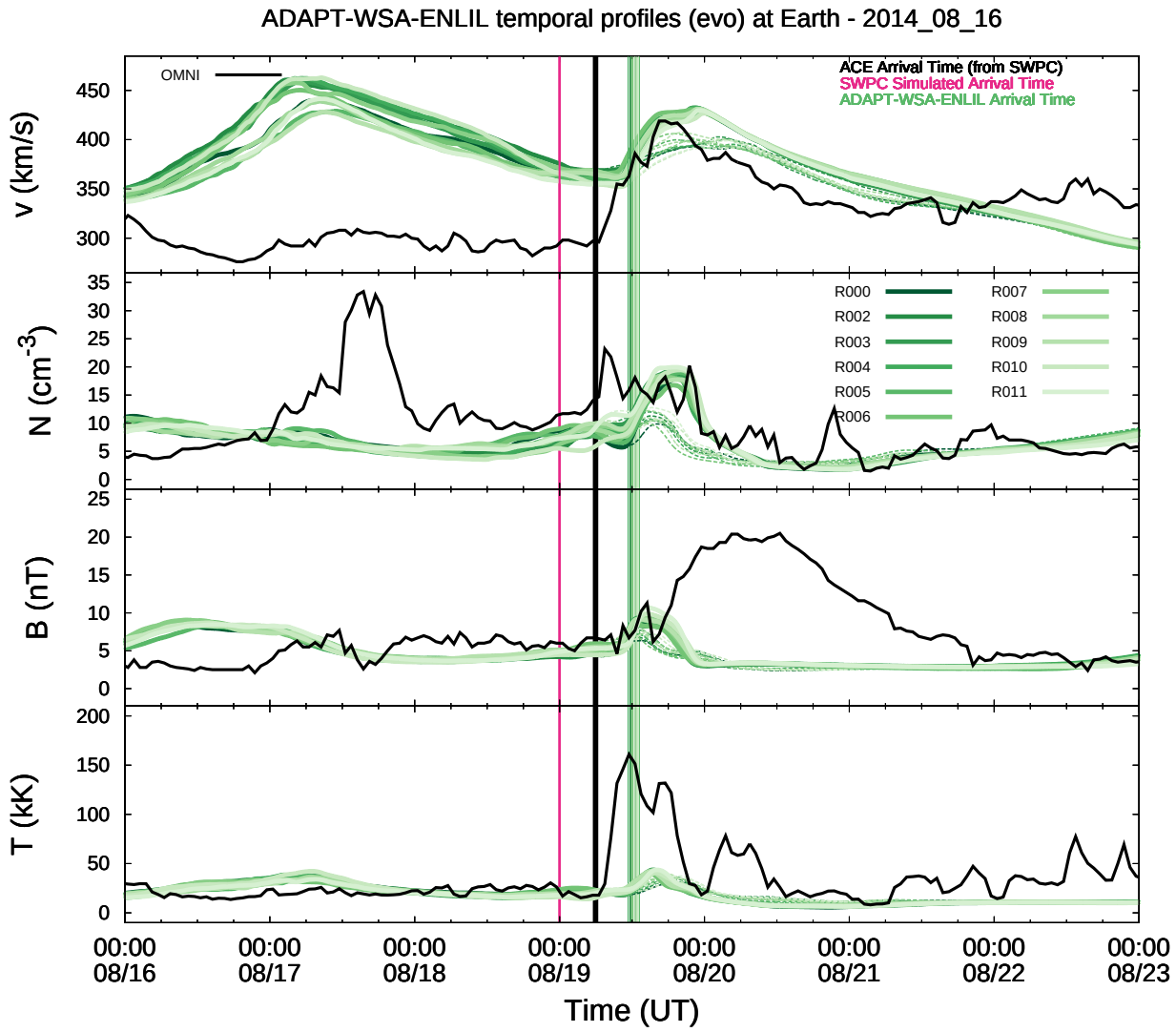

Figure 6. Time-series for 11 ADAPT-WSA-ENLIL simulations (shades of green; dashed lines show ambient simulation) for the 2014-08-15 CME, using a 2 hour ADAPT-WSA cadence at the ENLIL inner boundary.

models to test their performance using the same input parameters as the SWPC/CCMC project.

MLM thanks C. Verbeke, T. Jensen, B. Brown, P. Riley and the CME Arrival Time and Impact Working Team members for useful conversations.

REFERENCES

Brueckner, G. E., Howard, R. A., Koomen, M. J., et al. 1995, Sol. Phys., 162, 357 3
Domingo, V., Fleck, B., & Poland, A. I. 1995, Sol. Phys., 162, 1 3
Harvey, J. W., Hill, F., Hubbard, R. P., et al. 1996, Science, 272, 1284 1
Howard, R. A., Moses, J. D., Vourlidas, A., et al. 2008, Space Sci. Rev., 136, 67 3
Jolliffe, I., & Stephenson, D., eds. 2011, Forecast Verification: A Practioner's Guide in Atmospheric Science, 2nd edn. (New Jersey, USA: Wiley) 5
Kaiser, M. L., Kucera, T. A., Davila, J. M., et al. 2008, Space Sci. Rev., 136, 5 3
Stone, E. C., Frandsen, A. M., Mewaldt, R. A., et al. 1998, Space Sci. Rev., 86, 1 3
Verbeke, C., Mays, M. L., Temmer, M., et al. 2018, Space Weather, in revision 7.2
Wold, A. M., Mays, M. L., Taktakishvili, A., et al. 2018, Journal of Space Weather and Space Climate, 8, A17 5

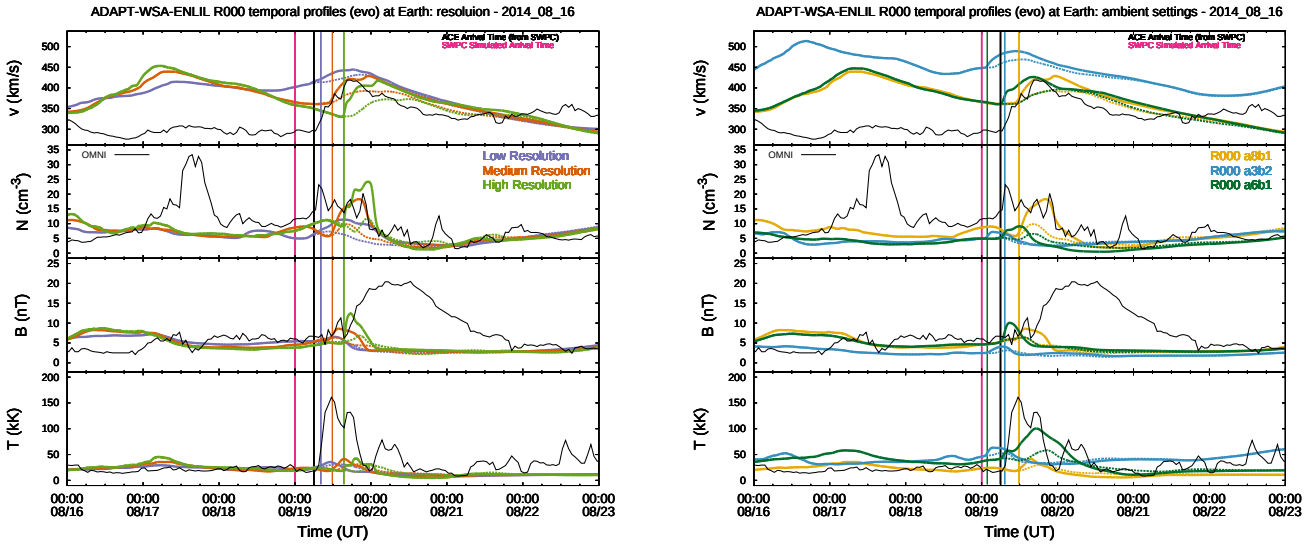

Figure 7. ADAPT realization R000. Left: time-series for low (purple), medium (orange) and high (green) resolution grids. Right: time-series for amb=a8b1 (yellow), amb=a3b2 (blue), and amb=a6b1 (green).

S1. SUPPLEMENTAL FIGURES & TABLES

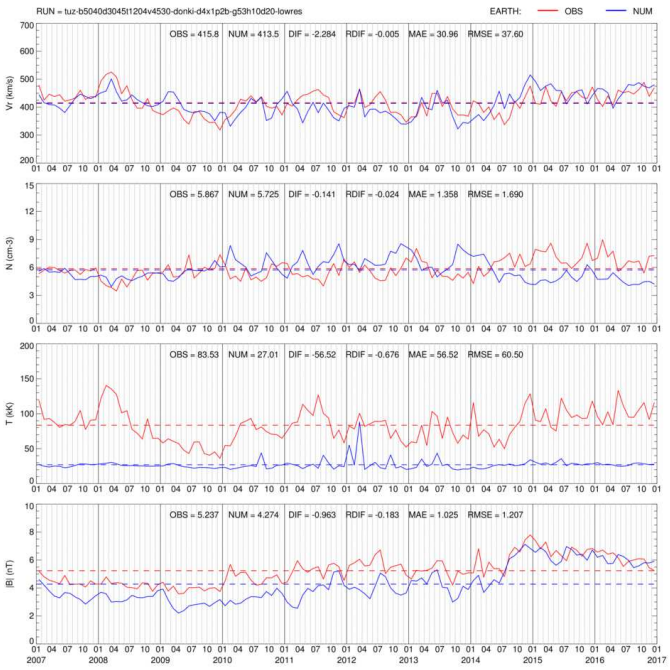

Figure S1. Comparison of WSA version 2.2 and ENLIL version 2.9f results when using zero-point corrected maps corobs=gongz. Also see Figure S2.

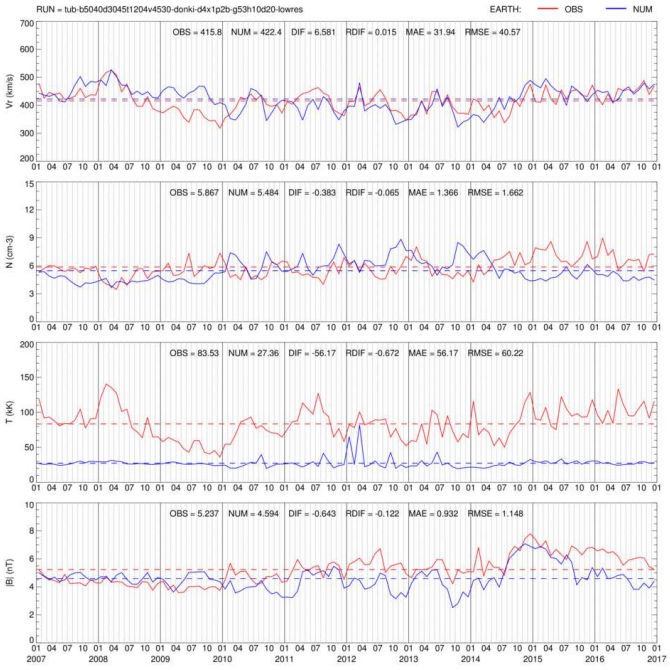

Figure S2. Comparison of WSA version 2.2 and ENLIL version 2.9f results when using zero-point uncorrected maps corobs=gongb, compare to Figure S1. Also see Figure 5 for more comparisons.

Table S1
Summary of CME observed and simulated arrival times along with the arrival time prediction errors.

Event #	CME Start Time [UT]	Arrival Time Observed [UT]	Arrival Time SWPC Operational [UT]	Δt (hours)	Arrival Time Single map [UT]	Δt (hours)	Arrival Time Time-dependent [UT]	Δt (hours)
1	2012-01-23 04:00	2012-01-24 14:31	2012-01-24 14:00	-0.52	2012-01-24T15:37:51	1.1	2012-01-24T18:48:26	4.28
2	2012-02-10 20:00	2012-02-14 07:03	2012-02-14 07:00	-0.05	2012-02-14T05:28:31	-1.58	2012-02-14T05:14:56	-1.82
3	2012-02-24 03:46	2012-02-26 20:58	2012-02-26 20:00	-0.97	2012-02-26T20:02:29	-0.93	2012-02-26T16:30:16	-4.47
4	2012-03-07 01:25	2012-03-08 10:45	2012-03-08 10:00	-0.75	2012-03-08T11:19:35	0.57	2012-03-08T11:05:11	0.33
5	2012-04-02 02:12	2012-04-05 20:03	2012-04-05 18:00	-2.05	2012-04-05T10:49:07	-9.23	2012-04-05T16:19:48	-3.73
6	2012-04-18 17:24	2012-04-21 09:25	2012-04-21 16:00	6.58	2012-04-21T16:07:17	6.7	2012-04-21T17:49:49	8.40
	2012-04-19 07:00	2012-04-23 02:27	2012-04-22 23:00	-3.45	2012-04-23T12:15:27	9.8	2012-04-23T15:25:18	12.97
7	2012-05-17 01:48	2012-05-20 01:36	2012-05-19 20:00	-5.60	2012-05-19T20:00:13	-5.6	2012-05-19T20:43:28	-4.88
8	2012-06-13 13:25	2012-06-16 09:01	2012-06-16 19:00	9.98	2012-06-16T20:18:50	11.28	2012-06-16T19:27:11	10.43
	2012-06-14 14:12	2012-06-16 19:31	2012-06-16 19:00	-0.52	2012-06-16T20:18:50	0.78	2012-06-16T19:27:11	-0.07
9	2012-07-12 16:36	2012-07-14 17:28	2012-07-14 12:00	-5.47	2012-07-14T13:44:20	-3.73	2012-07-14T13:49:00	-3.65
10	2012-07-28 21:12	2012-08-02 09:22	2012-08-02 08:00	-1.37	2012-08-02T07:07:50	-2.25	2012-08-02T07:05:06	-2.28
11	2012-08-31 20:00	2012-09-03 11:23	2012-09-03 17:00	5.62	2012-09-03T17:08:26	5.75	2012-09-03T17:42:28	6.32
12	2012-09-28 00:00	2012-09-30 22:13	2012-09-30 15:00	-7.22	2012-09-30T15:15:09	-6.97	2012-09-30T15:23:30	-6.83
13	2012-10-05 02:48	2012-10-08 04:31	2012-10-08 15:00	10.48	2012-10-08T16:29:55	11.97	2012-10-08T13:55:43	9.40
14	2012-10-27 12:48	2012-10-31 14:40	2012-10-31 21:00	6.33	2012-10-31T21:07:06	6.45	2012-10-31T23:43:00	9.05
15	2012-11-20 12:00	2012-11-23 21:12	2012-11-23 17:00	-4.20	2012-11-23T15:52:19	-5.33	2012-11-23T17:01:07	-4.18
16	2013-01-13 06:00	2013-01-16 23:00	2013-01-17 05:00	6.00	2013-01-17T04:37:15	5.62	2013-01-17T09:19:43	10.32
17	2013-01-13 06:00	2013-01-16 23:00	2013-01-17 05:00	6.00	2013-01-17T05:21:55	6.35	2013-01-17T09:16:23	10.27
	2013-01-15 08:48	2013-01-19 16:47	2013-01-19 16:00	-0.78	2013-01-19T14:30:34	-2.28	2013-01-19T14:54:34	-1.88
18	2013-03-12 10:36	2013-03-15 05:05	2013-03-15 02:00	-3.08	2013-03-15T00:40:35	-4.42	2013-03-15T00:26:21	-4.65
19	2013-03-15 07:12	2013-03-17 05:28	2013-03-16 21:00	-8.47	2013-03-16T21:05:52	-8.38	2013-03-16T19:43:45	-9.75
20	2013-04-11 07:24	2013-04-13 22:15	2013-04-13 21:00	-1.25	2013-04-13T20:48:51	-1.45	2013-04-13T21:09:16	-1.10
21	2013-04-26 18:24	2013-04-30 08:57	2013-04-30 09:00	0.05	2013-04-30T09:13:05	0.27	2013-04-30T12:00:00	3.05
22	2013-05-17 09:12	2013-05-19 22:21	2013-05-19 14:00	-8.35	2013-05-19T13:50:00	-8.52	2013-05-19T12:33:18	-9.80
23	2013-05-22 08:48	2013-05-24 17:35	2013-05-24 10:00	-7.58	2013-05-24T10:52:05	-6.72	2013-05-24T09:23:10	-8.20
24	2013-06-28 02:00	2013-06-30 10:40	2013-06-30 14:00	3.33	2013-06-30T14:27:34	3.78	2013-06-30T14:03:28	3.38
25	2013-07-06 19:36	2013-07-09 19:58	2013-07-09 23:00	3.03	2013-07-10T00:07:41	4.15	2013-07-10T03:40:04	7.70
26	2014-01-04 19:54	2014-01-07 14:25	2014-01-07 09:00	-5.42	2014-01-07T09:14:45	-5.18	2014-01-07T15:49:12	1.40
27	2014-01-07 18:24	2014-01-09 19:31	2014-01-09 08:00	-11.52	2014-01-09T09:48:29	-9.72	2014-01-09T09:48:08	-9.72
28	2014-03-23 03:36	2014-03-25 19:25	2014-03-26 06:00	10.58	2014-03-26T05:23:28	9.97	2014-03-26T10:24:05	14.98
29	2014-04-17 01:25	2014-04-20 10:24	2014-04-20 03:00	-7.40	2014-04-20T03:04:33	-7.33	2014-04-20T05:00:16	-5.40
	2014-04-17 12:48	2014-04-20 10:24	2014-04-20 14:00	3.60	2014-04-20T14:52:40	4.47	2014-04-20T15:30:23	5.10
30	2014-08-15 17:48	2014-08-19 05:58	2014-08-19 00:00	-5.97	2014-08-18T23:55:24	-6.05	2014-08-19T00:12:25	-5.77
31	2014-09-02 16:48	2014-09-06 14:19	2014-09-06 09:00	-5.32	2014-09-06T09:13:39	-5.1	2014-09-06T04:38:16	-9.68
32	2014-09-09 00:06	2014-09-11 22:58	2014-09-12 00:00	1.03	2014-09-12T01:52:57	2.9	2014-09-12T00:56:41	1.97
33	2014-09-09 00:06	2014-09-11 22:58	2014-09-12 00:00	1.03	2014-09-11T23:05:17	0.12	2014-09-12T00:55:56	1.95
	2014-09-10 18:00	2014-09-12 15:30	2014-09-12 14:00	-1.50	2014-09-12T15:19:45	-0.18	2014-09-12T14:13:39	-1.28

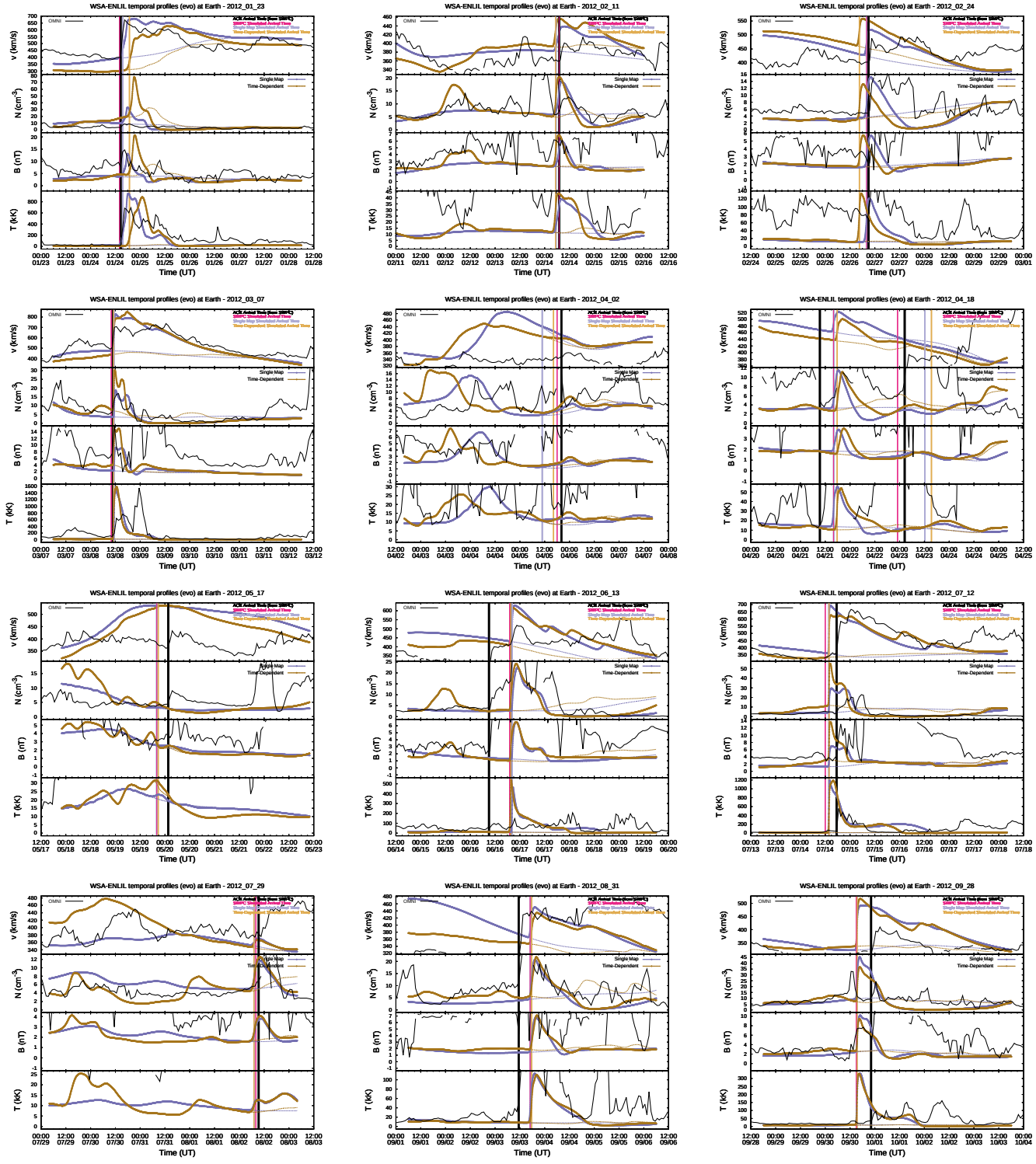

Figure S3. Simulation results for the benchmark single map-driven runs (Section 2, task a) and time-dependent map-driven runs (Section 2, task b) for all 33 events. The simulated benchmark time-series are plotted in purple, time-dependent runs in brown, and the hourly averaged OMNI data is plotted in black. The green vertical lines show the SWPC provided simulated arrival times, the purple vertical lines show those detected by the dynamic pressure algorithm, and the black vertical lines show the ACE observed arrival times.

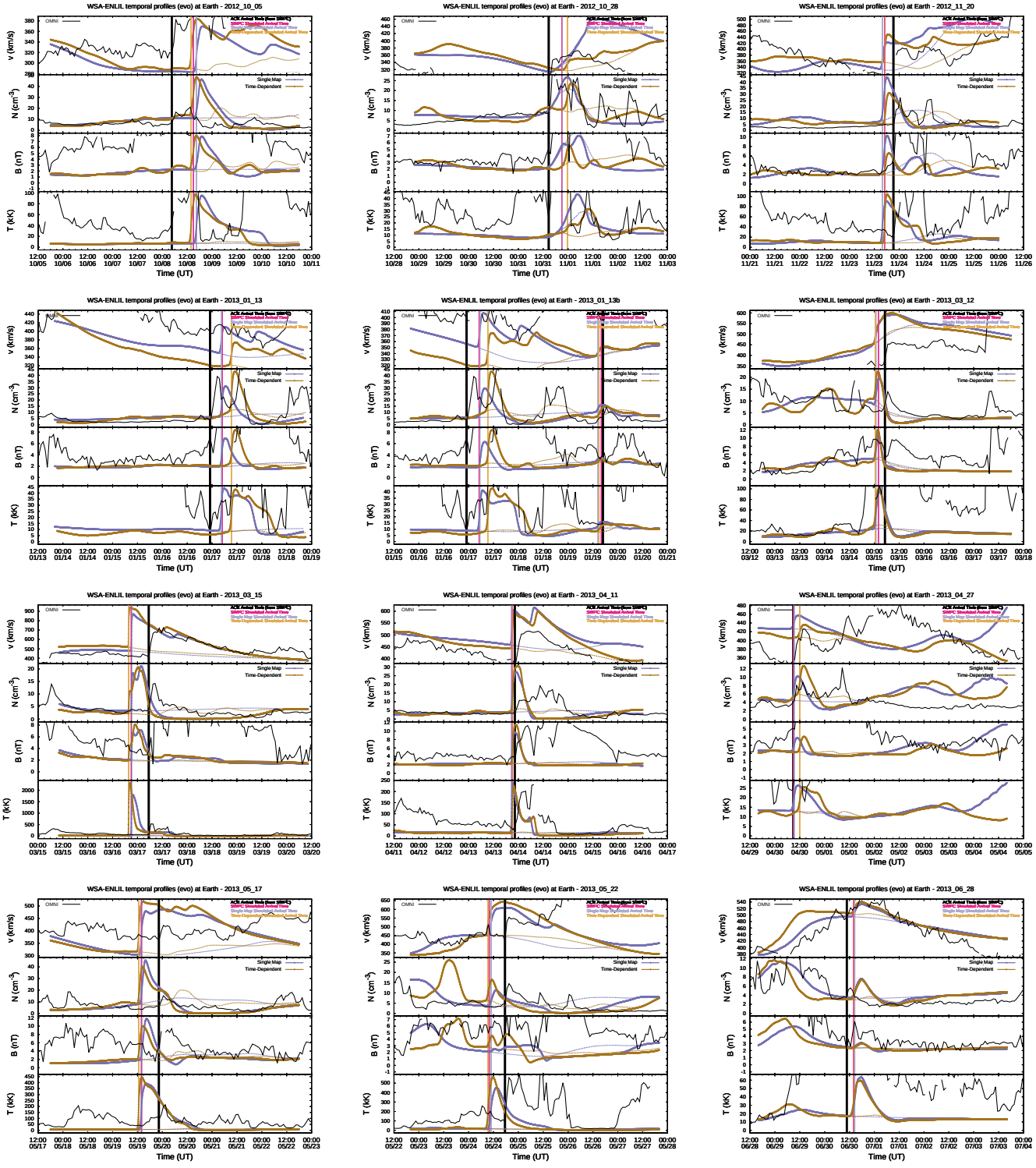

Figure S4. Figure S3 continued.

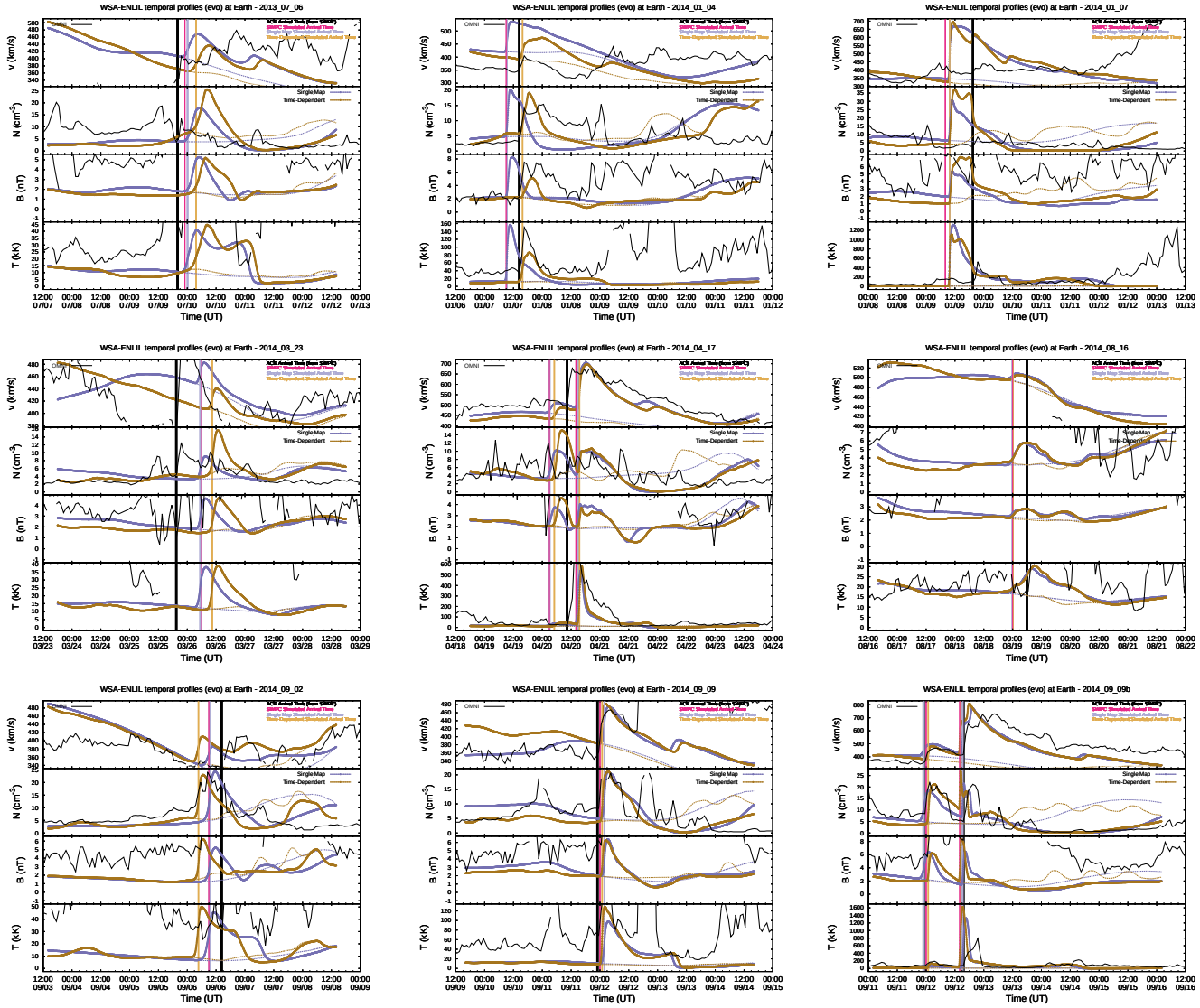

Figure S5. Figure S4 continued.

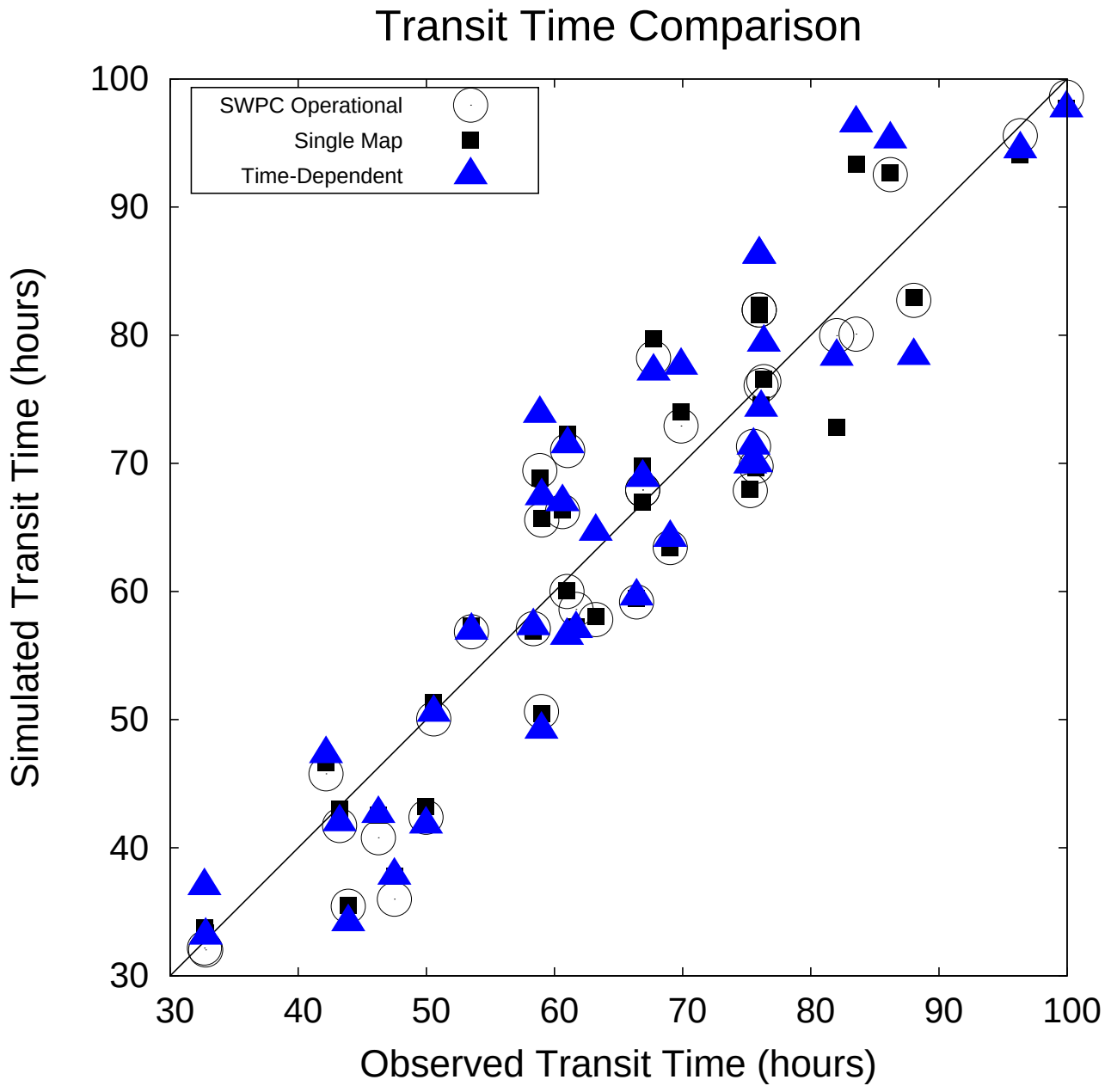

Figure S6. Predicted vs observed CME transit times for each simulation setting.

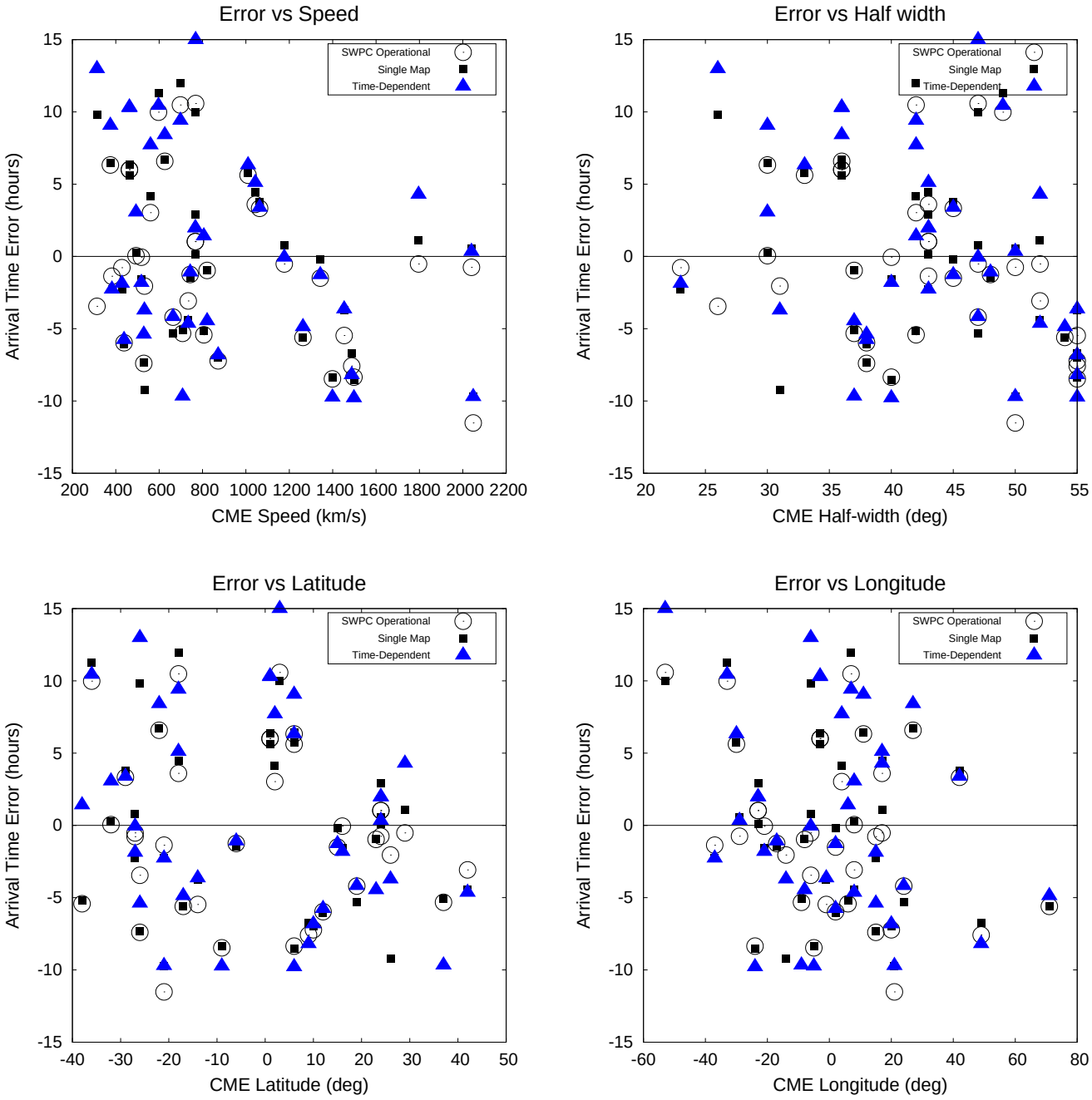

Figure S7. Dependence of the arrival time error on CME input parameters.

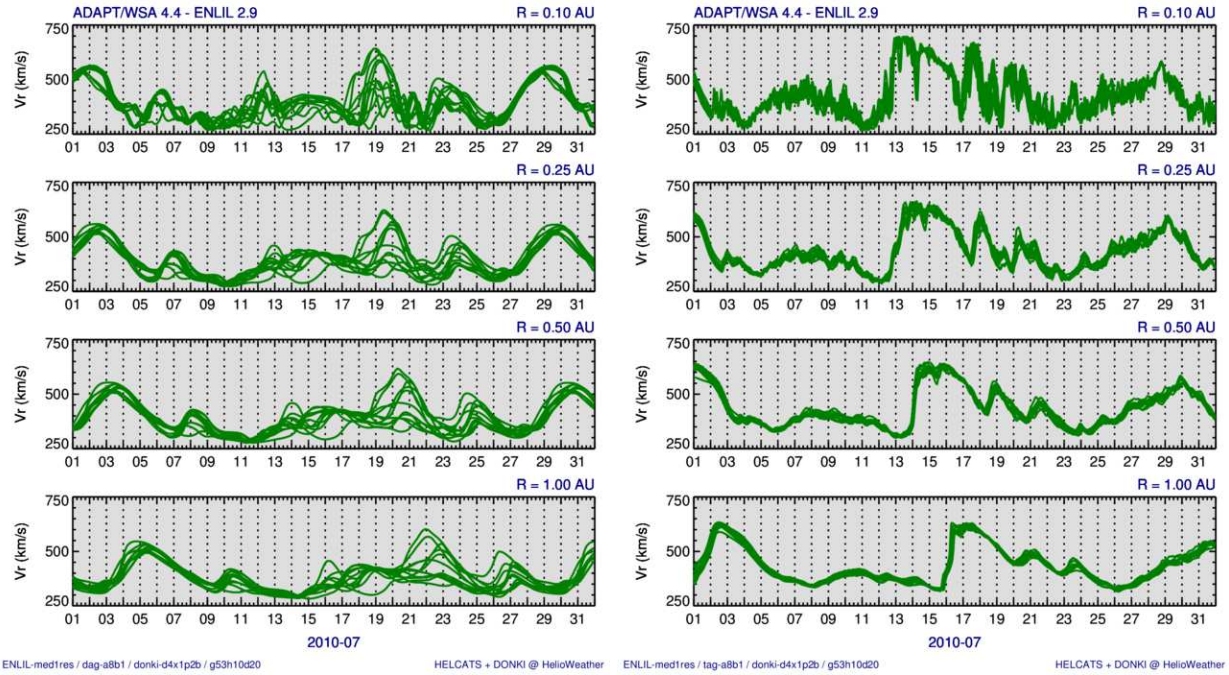

Figure S8. Twelve ADAPT-WSA-ENLIL radial velocity model outputs at different radial distances (0.1, 0.25, 0.5, 1 AU) for July 2010 using a 24 hour ADAPT-WSA input cadence (left) or 2 hour input cadence (right).

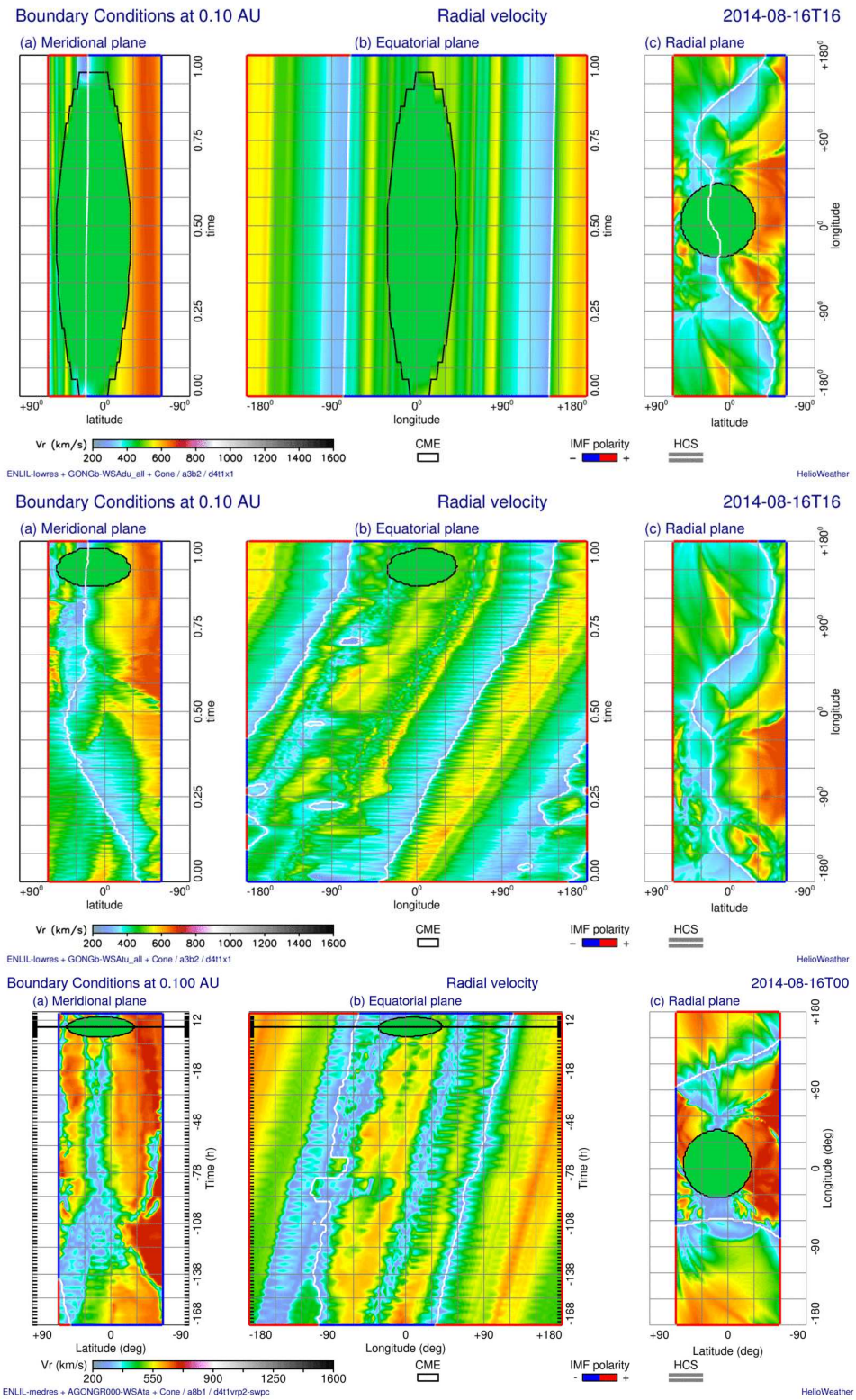

Figure S9. Panel (b): equatorial slices of the ENLIL inner boundary as a function of time. The top figure shows the single map-driven case (cormode=single), the middle figure shows the time-dependent map-driven case (cormode=multi), and the bottom figure shows the time-dependent ADAPT map-driven case for realization R000 (cormode=multi). This figure illustrates that for the single map-driven case (bottom), the background map is fixed and rotated through time during CME insertion, whereas for the time-dependent sequence of maps (middle, bottom), the CME is inserted into a clearly time-varying background.