Summary International CCMC-LWS Working Meeting: Assessing Space Weather Understanding and Applications

April 3 - 7, 2017, Cape Canaveral, Florida

Please fill out the opinion survey: https://goo.gl/z8EnwK

Working Team Focus and Goals

This team will evaluate how well different models/techniques can predict CME arrival time and impact for a set of historical events, with open communication with the community. The work is complementary to the CME Scoreboard activity whose goal is collect and display real-time CME predictions and facilitate the validation of real-time predictions.

- ✳ Evaluate where we stand with CME arrival time and impact prediction
- ✳ Establish metrics agreed upon by the community that address both science and user needs
- ✳ Provide a benchmark against which future models can be assessed against

Working Team Deliverables

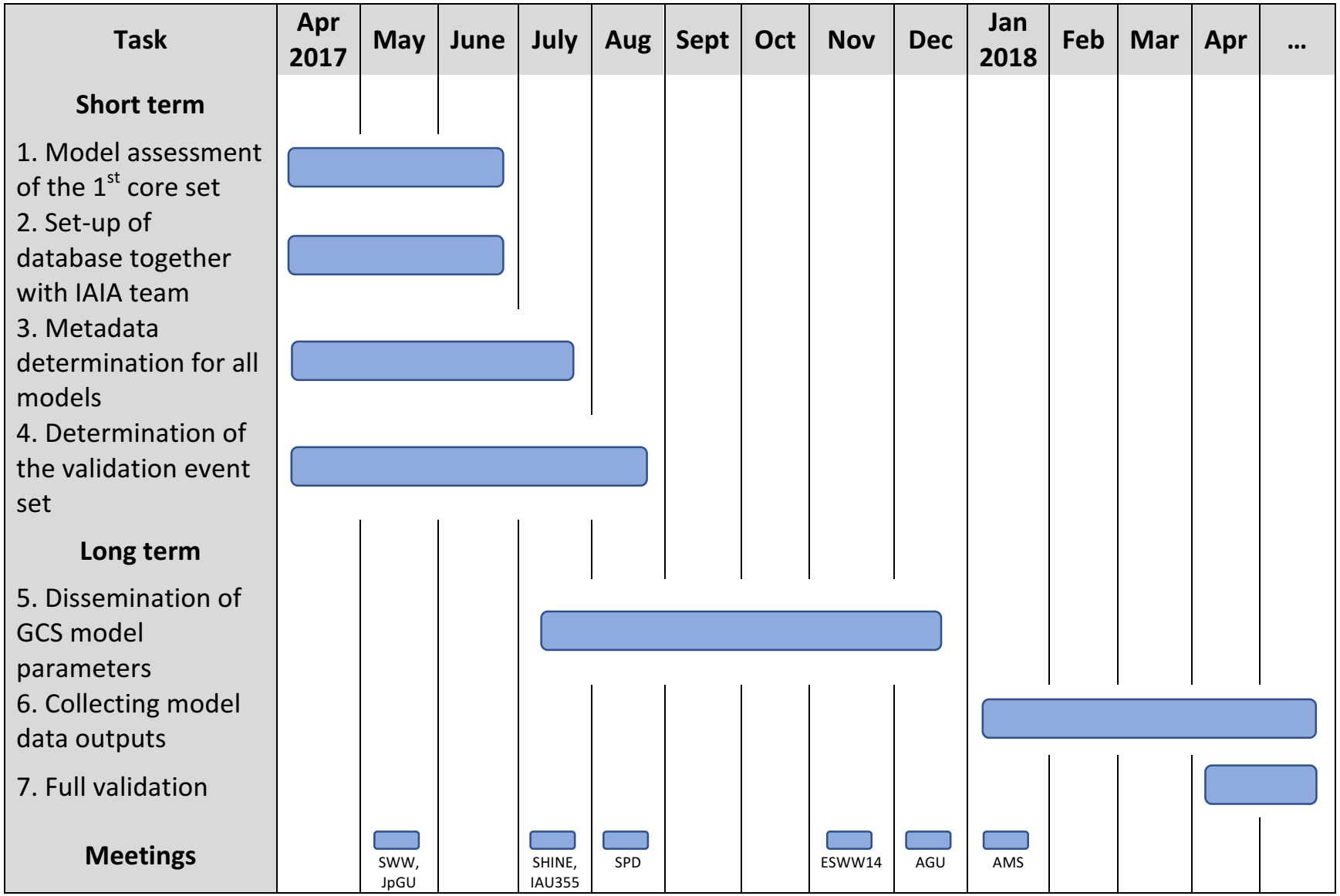
- ✳ Catalog of metrics and how they relate to user needs and validation needs.
- ✳ Model assessments with selected metrics for selected time intervals.
- ✳ Online database of model inputs, outputs, and observations.
- ✳ Publication describing model assessment results summarizing where we stand with CME arrival time and impact prediction.

Upcoming meetings for team and community interaction:

Participants can meet informally at upcoming meetings. There are plans for a session at AGU jointly with IMF Bz and 3D CME kinematics teams, and also for a topical discussion at ESWW14.

Name	Time	Place	Deadline session?	Deadline abstract?
Space weather week	1-5 May 2017	Broomfield, CO, USA	NA	Passed
JpGU-AGU	20-25 May 2017	Chiba Japan	Passed	Passed
IAU symposium 335	July 17-21 2017	Exeter, UK	Passed	Passed
SHINE	July 24-28 2017	Saint-Sauveur, Canada	Passed	June 23, 2017
SPD	21-25 Aug 2017	Portland OR, USA		May 30,2017
ESWW14	Nov27-Dec1 2017	Belgium	Sessions-Passed Topical discussion: May 10, 2017	May 31, 2017 posters Nov 1, 2017
AGU	11-15 Dec 2017	New Orleans, USA	April 19, 2017	August 7, 2017
AMS	7-11 January 2018	Austin, TX, USA	May 1, 2017	August 1, 2017

Future tasks and planning:



Event selection and model input:

During the meeting, we have discussed the event selection. We agreed to take a set of 100 events for statistical significance. Both users as well as scientists agreed that all types of CMEs should be taken into account, but we shall flag each of the CMEs into their corresponding category for later use and statistics.

For the CME arrival time and plasma parameters from observations, the following has been proposed:

- CME arrival time will be taken from current existing CME catalogues. For the cases where the CME can be found in multiple catalogues, a random catalogue will be chosen.
- 1 hour averaged OMNI data will be used for solar wind impact parameters such as the peak of the density/velocity.

As our goal is to see where each model stands and what arrival and solar wind parameters at impact of your model are underperforming, and which are performing well, so that we can advance and improve our models, we would like to fix the model input for the 100 events. We have come up with the following situation:

- Each model will provide results for the 100 events for the fixed input parameters that are provided by the 3D CME kinematics team as well as fixed magnetogram. The 3D CME kinematics team will provide the GCS model inputs. In case your model does not use such input, or slightly adjusted input, an "in-between" program/algorithm may be used to process the provided input parameters. This will only be acceptable if all 100 parameters are processed in the exact same manner.
- For those models that cannot provide results for all 100 events, a subset will be provided.
- Optionally, modelers can also submit the 100 events for their best parameters.
- There will be an extra (smaller) set with events that have CME observations at other 1AU spacecraft. It is not necessary for the models to perform their validation on this set. However, it is highly appreciated as it will leave us with a quantification of how well we can perform compared to observations at Earth.

In agreement with the IMF Bz team, we will have some event overlap so that we can lower the burden for those models that are providing data/model outputs for both teams.

Open discussion on events:

- Agreement on the above event selection and model input process
- What to do with multiple spacecraft events?
- What solar cycles do we want to consider for the event selection?
- How much does the resolution of the model influence the results?
- Is there a preference towards having the events released in smaller batches or all at once?

Metrics and skill scores:

The first discussion item that came up during the working meeting was on how to define a hit. It seems that modelers and users have different needs towards which interval is more suitable to approach. Therefore, we have agreed that, once the model inputs are received, we will compute skill scores using different hit definition intervals: 3h, 6h, 12h, 18h, 24h and 36h.

For each of these hit interval definitions, we will set up contingency tables. Scores include: Hit rate (POD), False alarm rate, False alarm ratio, Bias, Accuracy, Threat score, Base rate, Proportion correct, HK, HSS.

In addition to the contingency table, we will also determine a set of errors: RMSE (Root Mean Square Error), MAE (Mean Absolute Error), and the ME (Mean Error). The mean error evaluates how the model performs regarding early or late arrivals etc.
For the users, if the model arrival time includes a confidence interval, a model confidence interval contingency table will be computed. In this case, a hit will be an observed arrival within the model's confidence interval.

For the CME plasma parameters at arrival, we will make use of correlation coefficients. This allows us to determine the performance of the model output parameters. In Figure 1 an example can be found.

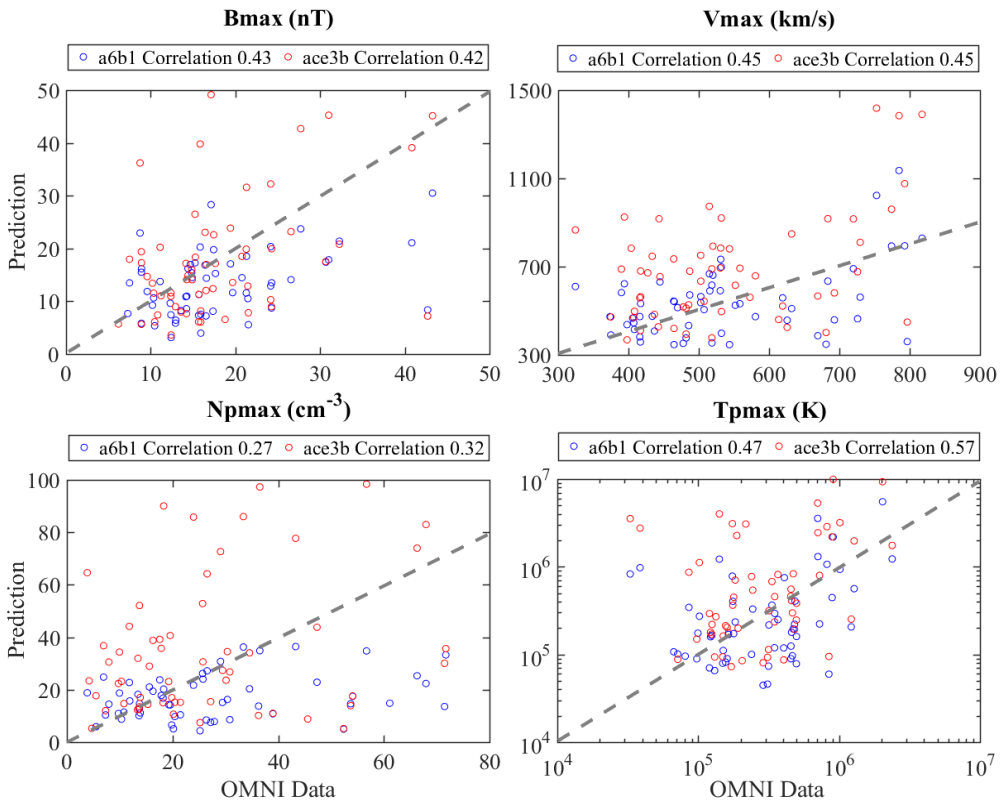


Figure 1: Correlation coefficients for CME impact parameters (courtesy: Lan Jian).

Open discussion on metrics and skill scores:

- How will each model define the modeled CME arrival time? Use the same algorithm for similar models? Model output submitter's decision? Similar question for the duration of the event. The metadata will keep track of human vs. algorithm identified model arrivals.
- How do we want to compare the time series of the modeled events to the observed time series? The IMF Bz team discussed time shifting the model time series to the observed time series. However, this can be done multiple ways:
    - Shift towards same arrival time
    - Shift towards density/velocity peak
    - Shift so that the performance of model is best (IMF Bz team choice)

  This is still under discussion.
- Is there a need for a combined skill score for user that can summarize overall model performance?

Background solar wind influence

We have discussed the influence of the background solar wind influence but we have not yet found a way to quantify this influence. A plot shown for discussion using ENLIL results is shown in Figure 2. It seems that the background solar wind does not have an obvious influence on the arrival times. We will need further investigation for this. This item is still open for discussion.

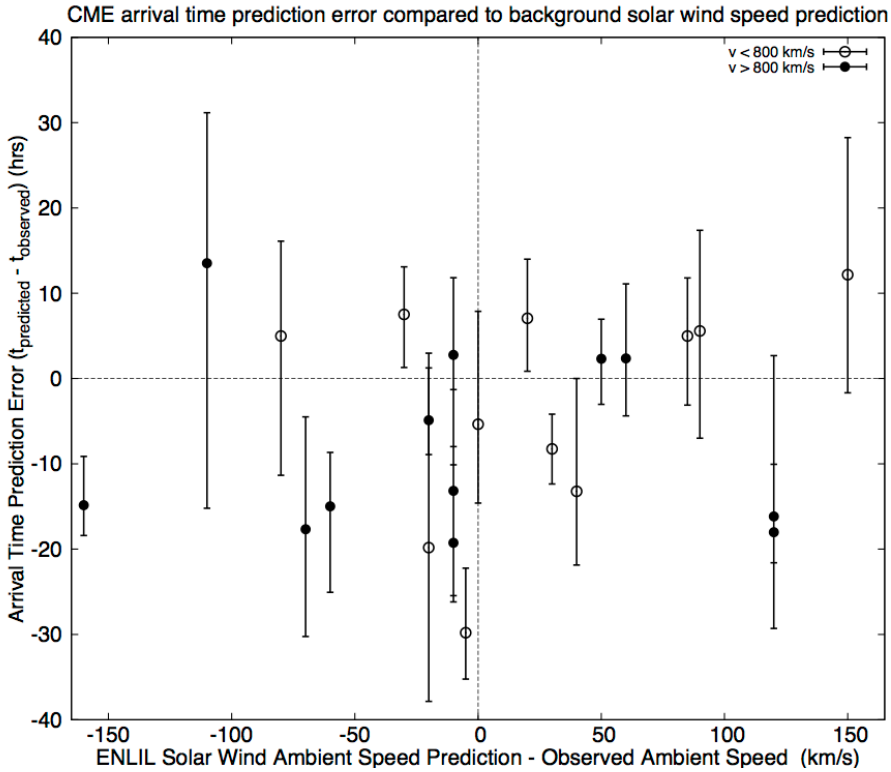


Figure 2: Comparison of arrival time errors with errors in ambient solar wind prediction 12 hours prior to ICME arrival.

Other/various:

During the discussions, the need for an enhancement to the CME scoreboard came forward.

- Implementing a system where CME misses (ICME observed but not predicted) can be added so that users can have feedback on why the CME was not forecast in real-time, as well as having a database with CME misses.
- In addition, for each "active CME" participants can submit a prediction of "no arrival" (that the CME will not arrive).
- CME scoreboard is currently working on an XML file for automated model submissions