

Draft Findings under the iCCMC assessment for Bz at 1AU.

Executive Summary

NASA's CCMC Validation Activity

International Forum for Space Weather Capabilities Assessment

Scientific Domain

Heliosphere

Focused Evaluation Topics

IMF Bz at L1

Title description:

To create a community-agreed selection of events and metrics, that all current and future models should test their magnetic field forecasting capabilities.

Abstract description:

In this topic the community will focus on forecasting the magnetic structure of interplanetary CMEs and the ambient solar wind upstream of Earth. This group intends to open communication with the community in order to agree upon a standardised process by which all current and future models can be compared under an unbiased test. Current models will provide the initial set of forecasting skills, with the longer-term goal of providing a standardised test procedure which future model improvements can follow. This procedure is intended to provide concrete requirements to progress a scientific model along the Application Readiness Levels (ARLs) and into an operational setting. The conversation and scientific rationale behind all decisions will be recorded to facilitate future ARL procedures.

Models:

Current list of models incorporated in our discussions are:

Data driven -

1. Bz4Cast (N. Savani)
2. Helicity-CME, H-CME (S. Patsourakos)
3. A. Rouillard model

Numerical MHD simulation -

4. SUSANOO (D. Shiota)
5. EUHFORIA (S. Poedts)

Recommendation algos -

6. ProjectZed (P. Riley)

Purpose of Draft Findings:

We invite the wider community to provide further insight that would benefit the final determination of evaluation criteria especially in those areas that remain outstanding. All new ideas are welcome, as well as additional suggestions on current evaluation themes.

A small team of model developers and end users (SWPC and UKMO) were selected to 'seed' an initial direction for further discussion by the wider community.

The topics listed below are the themes under which initial discussions were made. A few of the groups initial findings and suggestions are also provisionally provided.

Subject matter themes:

1. Background Solar Wind Conditions
2. Core event selection
3. Magnetic What?
4. B Magnitude threshold
5. Time resolution

1. Background SW Conditions

1. Evaluation at a 1-minute resolution is probably required. Pragmatically, as most inputs into magnetospheric models are at 1min resolution, we expect a validation metric at this level is needed. We also expect most models will perform poorly under this criterion.
2. Datapoint to datapoint (dp2dp) evaluations should be included in a final set of metrics, even though they might be misleading if not properly assessed.
3. Actionable intelligence can be estimated by a forecast that provides accurate |B| and fluctuation intensity, regardless of dp2dp phase. How this is evaluated remains outstanding.
4. "Event-based" metrics should also be performed. Such as for High speed stream. How this is evaluated remains outstanding.

2. Core event selection

1. As a process of advancing models from research into the operational domain (R2O), the suggestion is to include 3 phases of metric evaluation.
 - a. A small core selection of 'ideal case studies'. – ~10 events so that the initial hurdle for future models remains low.
 - b. A full collection of realistic events. – ~50 events that include the full range of 'complicated' behaviour observed in the heliosphere
 - c. Blind selection. – a final group of events that also covers the full domain of observable CME events. This group of events will remain blind to model developers until their models have passed all previous ARL requirements. Models will be required to test forecasting skill in a virtual real-time environment.
2. The SCOSTEP/ ISEST event list has been identified a useful source of events from which 'textbook' and other 'non-ideal' events may be selected. Final selection of events remains outstanding.
3. Overlapping CME events with iCCMC Arrival Time Topic will be of added benefit to the wider community. Further cross-talk remains active.

3. Magnetic What?

1. Cartesian coordinate system (opposed to spherical) of the magnetic vectors has shown significant support from both model developers and end users.
2. Distilling the vector direction into a single skill metric is not a necessary goal for end users (SWPC or UKMO). Best to approach the problem with multiple metrics that can, at a later stage, be condensed into a meaningful way by the operators.
3. UKMO would prefer metrics performed in GSE coordinates.
4. RTN system (locally Cartesian) would provide a more universal metric that remains applicable to all future interplanetary endeavours. While this is not ideal from an Earth Space weather perspective, it might be manageable for operators such as UKMO. A louder preference from the wider community between RTN and GSE would assist with the final decision.

4. B Magnitude threshold

1. A strong preference has been shown for multiple thresholds. This approach also assists with further activity into ensemble research.
2. A current procedure of action at both SWPC and UKMO is triggered for $B_z < -5nT$ sustained for more than 30min. This provides an initial benchmark of interest.
3. Further intensity thresholds and for what time duration remains outstanding.

5. Time resolution

1. Metrics developed at 1 minute resolution is of benefit to a variety of users – e.g. magnetospheric and GIC applications.
2. 1min, 5min, and 1 hour from the OMNIweb data rates are currently considered for further evaluation.
3. Keeping to the OMNI/ COHWeb data standards enables for the widest possible community to participate using standardised data inputs. – Suggestions have been made to avoid getting bogged down into such prolonged scientific discussions of how best to average the data at this early stage.
4. For the sheath region downstream of a CME-driven shock, concerns have been raised that 1 hour resolution would not capture desired variations and 5 minute would include too much. How best to evaluate the sheath remains outstanding.
5. Do we need different resolutions for different applications and end users? If so, which communities and how many are there. This question remains outstanding.

6. Validation Metric

1. General consensus is that a broader collection of skill metrics are required moving forward. We should aspire to develop metrics that are more comprehensive than the current 'Probability of Detection' used by SWPC.
2. Very strong advocacy for reporting confidence (uncertainty) levels. There is strong support on both scientific community and by end users.
3. However, an outstanding issue remains on how to incentivise the scientific community to increase efforts towards more accurate estimates of the confidence levels. I.e. "There's a big difference in usability between a forecast of 'danger' with unknown confidence, a forecast of 'danger with 60% probability'".
4. Traditional weather forecasting has significant experience which should be leveraged wherever possible. "Dysfunctional skills are taught by using simplistic means [e.g. GPRA] to measure value". It remains outstanding on how this community aims to ensure lessons learnt by weather forecasting will be implemented.