The First Magnetospheric Space Weather Metrics Challenge

A Report

Michael Hesse

CCMC

NASA/GSFC

GEM

6/19/01
Background: History

• NSWP Iplan calls for evaluation of models

• Metrics study (R. Wolf and T. Fuller-Rowell, chairs) defines parameters

• NSF/R. Behnke calls for evaluation of magnetospheric models

• NSF tasks GGCM Steering Committee with implementation of first metric

• G. Siscoe leads implementation

• CCMC performs evaluation and reports results at SWW

• Additional participants now included in report to GEM
Metrics for This Study

**Goal:** compare velocity measurements from DMSP F13 with (ExB drift) velocities from models

Precisely:
If $v_y$ is the DMSP cross-track component of the perpendicular (to B, ion) flow velocity, and $v_{ym}$ is the corresponding quantity, derived from model m, then the deviation or metrics $D_m$

$$D_m = \frac{1}{N} \sum_{i=1}^{N} |v_{y,i} - v_{ym,i}|$$

Skill Score:
If $D_g$ is the deviation for a given (e.g., commonly used) model g, the skill score of model m relative to model g is

$$S_{m,g} = 1 - \frac{D_m}{D_g}$$

Note: S<0 possible!

Here: g is the model, which predicts $v_y=0$ (!)
Event Selection

Based on quality of solar-wind and DMSP F13 data, and consensus

Period 1: 16-17 April 1999, Days of year = 106-107 (prime period)
Period 2: 10-11 December 1998, Days of year = 344-345
Period 3: 05-06 November 1998, Days of year = 309-310

Data from ACE, Geotail, IMP8, Interball Tail, and Wind, collected by M. Heinemann

DMSP F13 passes
  Period 1: 33
  Period 2: 40
  Period 3: 37

DMSP data provided by M. Hairston
Participants/POCs

B. Hausman/RiceU: Toffoletto-Hill Model
J. Raeder/UCLA: UCLA MHD model
A. Ridley/Umich: BATSRUS MHD model
S. Slinker/NRL: LFM MHD model
R. Winglee/Uwash: Winglee MHD model
D. Weimer/MRC: Weimer PC model
D. Weimer/MRC-ISM: ISM MHD model
M. Wiltberger/Dartmouth: LFM model
CCMC (blind test only): BATSRUS MHD model
## Level of Participation

<table>
<thead>
<tr>
<th>Participant/Period</th>
<th>November passes</th>
<th>December passes</th>
<th>April passes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hausman</td>
<td>12</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Raeder</td>
<td>34</td>
<td>40</td>
<td>31</td>
</tr>
<tr>
<td>Ridley</td>
<td>37</td>
<td>40</td>
<td>33(twice)</td>
</tr>
<tr>
<td>Slinker</td>
<td>13 (also F14)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Winglee</td>
<td>-</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>CCMC</td>
<td>-</td>
<td>-</td>
<td>17</td>
</tr>
<tr>
<td>Weimer/PC model</td>
<td>22</td>
<td>40</td>
<td>26</td>
</tr>
<tr>
<td>Weimer/ISM</td>
<td>-</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>Wiltberger</td>
<td>-</td>
<td>19</td>
<td>-</td>
</tr>
</tbody>
</table>
trend for lower skill at later times?
Skill Scores November 98

- Slinker/skill
- Ridley/skill
- Raeder/skill
- Weimer/skill
- Rice/skill

Event # quality has same trend for all models
Skill Scores December 98

The diagram shows a scatter plot of skill scores for different individuals over various events. The x-axis represents the event number, ranging from 0 to 40, and the y-axis represents the skill score, ranging from -1 to 0.6. Different colors and markers are used to distinguish between individuals:

- Orange: Winglee/skill
- Blue: Ridley/skill
- Red: Raeder/skill
- Green: Weimer/skill
- Light green: Wiltberger/skill
- Light gray: Wiltberger/skill

The plot visually represents the performance trends and variations over time for these individuals.
A Few Select Passes

• High skill

• Low skill

• Medium skill 1

• Medium skill 2
A Particularly Good Case

models appear to match smoothed $v_y$ (100s) best
A Particularly “Bad” Case

DMSP data-model comparison

Spatial structure details not reproduced
Better match with smoothed $v_y$

Raeder: -0.19
Ridley: -0.05
Ridley/td: 0.0
CCMC: -0.43
Weimer: 0.16
Medium Skill Case 1
DMSP data-model comparison

Spatial structure details not reproduced

More detail (Rice, Slinker) does not necessarily lead to higher skill, although it is desirable.
Medium Skill Case 2

DMSP data-model comparison

Raeder: 0.38
Ridley: 0.34
Ridley/td: 0.53
Winglee: 0.35
Weimer: 0.57
CCMC: 0.17

Sharp boundaries not reproduced
Magnitudes are not too bad
April Period: ACE data

Lower model performance related to

- Large $B_z$?
- Large $B_y$? ←
- Density fluctuations?
Lessons learned

This was an interesting, very successful study

It went very well

For future metrics studies, we should
- precisely define the model data format early
- precisely define the relevant time intervals early
- define procedure of handling bad data points
- define procedure for comparing to runs w. bad data points
- precisely define who is responsible for what
- perform blind studies (my suggestion)
• There was no pass modeled by all participants
• All models are performing well, and better than v=0 prediction
• The same model can produce very different quality results for different passes
• There is no general trend for one model doing better under one condition and worse than another under a different condition
• It is very difficult to find passes where all models did well or all did badly
• All models show the same quality trends
Summary/2

• All models are roughly equivalent (bad statistics in some cases)

  The Michigan model provides roughly the same skill whether time dependent or not

  “Relatively simple” models (Weimer, Toffoletto-Hill) and the MHD codes provide similar quality results

• Models appear better at predicting flow magnitude than details such as flow reversal boundaries

• This was a great step in the right direction

• The results could be used for (friendly) in-depth studies and comparisons to improve overall skill scores (even further)