Geospace Model Workshop
April 25, 2011 Boulder, CO

Today’s Focus
• Metrics
• Selection Considerations
• Plans

Protecting Power Grids (and other services)

Howard Singer and Terry Onsager
NOAA Space Weather Prediction Center
Masha Kuznetsova, Antti Pulkkinen, Lutz Rastaetter
NASA Community Coordinated Modeling Center

Safeguarding Our Nation’s Advanced Technologies
Geospace Model Project Goals

- **Goal:** Evaluation of Geospace prediction models to determine which model or models should begin transition to operations process beginning January 2012.
- **Focus:** Models that can predict regional geomagnetic activity
- **Process:** CCMC leads evaluation; Build on GEM Storm Challenge; Establish partnerships; Select metrics; Conduct evaluation
- **Community Discussions:** GEM, AGU, and CCMC Meetings; Geomagnetic activity products documents circulated, Geospace Model Validation Workshop…
Models at CCMC Participating in Geospace Evaluation

• MHD Models:

• Space Weather Modeling Framework (SWMF) - U. of Michigan (delivered to CCMC)

• The Open Geospace General Circulation Model (Open GGCM) - University of New Hampshire (delivered to CCMC)

• Coupled Magnetosphere-Ionosphere-Thermosphere (CMIT) - BU CISM, Dartmouth, NCAR (delivered to CCMC)

• Grand Unified Magnetosphere-Ionosphere Coupling Simulation (GUMICS) - Finnish Meteorological Institute (not yet parallelized or ready for full evaluation, but showing progress)

• Empirical Models

• Weimer Empirical Model, Va. Tech (delivered to CCMC/may update)

• Weigel Empirical Model, George Mason (delivered to CCMC)
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Speaker</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>0900 – 0915</td>
<td><strong>Introduction and Meeting Goals</strong></td>
<td>Singer and Kuznetsova</td>
<td>Millennium Harvest House Hotel</td>
</tr>
<tr>
<td>0915 – 0945</td>
<td><strong>Model Delivery to CCMC—Status, Validation, and Related Activities</strong></td>
<td>Kuznetsova</td>
<td>Boulder, CO</td>
</tr>
<tr>
<td></td>
<td>Delivered Model Configurations and discussion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0945 – 1000</td>
<td><strong>Open-GGCM</strong></td>
<td>Raeder</td>
<td>Boulder, CO</td>
</tr>
<tr>
<td>1000 – 1015</td>
<td><strong>CMIT</strong></td>
<td>Wiltberger</td>
<td>Boulder, CO</td>
</tr>
<tr>
<td>1015 – 1030</td>
<td><strong>SWMF</strong></td>
<td>Ridley</td>
<td>Boulder, CO</td>
</tr>
<tr>
<td>1030 – 1045</td>
<td>Coffee Break</td>
<td></td>
<td>Boulder, CO</td>
</tr>
<tr>
<td>1045 – 1100</td>
<td><strong>GUMICS</strong></td>
<td>Palmroth (presented by</td>
<td>Boulder, CO</td>
</tr>
<tr>
<td>1100 – 1115</td>
<td><strong>Weigel Empirical</strong></td>
<td>Weigel (presented by</td>
<td>Boulder, CO</td>
</tr>
<tr>
<td>1115 – 1130</td>
<td><strong>Weimer Empirical</strong></td>
<td>Weimer (presented by</td>
<td>Boulder, CO</td>
</tr>
<tr>
<td>1130 – 1200</td>
<td><strong>Metrics for CCMC Test Runs, Test Run results, and Discussion</strong></td>
<td>(Singer and Pulkkinen)</td>
<td>Boulder, CO</td>
</tr>
<tr>
<td>1200 – 1330</td>
<td>Lunch</td>
<td></td>
<td>Boulder, CO</td>
</tr>
<tr>
<td>1330 – 1400</td>
<td><strong>Test Runs – continue discussion</strong></td>
<td>(Singer and Pulkkinen)</td>
<td>Boulder, CO</td>
</tr>
<tr>
<td>1400 – 1500</td>
<td><strong>Proposed Metrics</strong></td>
<td>Singer (All)</td>
<td>Boulder, CO</td>
</tr>
<tr>
<td>1500 – 1600</td>
<td><strong>Selection Considerations</strong></td>
<td>Onsager (All)</td>
<td>Boulder, CO</td>
</tr>
<tr>
<td>1600-1700</td>
<td><strong>Next Steps</strong></td>
<td>All</td>
<td>Boulder, CO</td>
</tr>
</tbody>
</table>
Geospace Model - Metric Tests

• To better design the final metrics that will be used for model evaluation, to establish experience with the metrics for discussions with modelers, and to explore how to best communicate selected metrics to the CCMC carrying out all simulations and the metrics calculations:
  • SWPC and CCMC have selected test metrics that have been run at CCMC on the outputs from the GEM Challenge studies (Pulkkinen et al., 2010; Rastaetter et al., 2010).

• \( \text{db/dt} \), i.e. the time derivative of the horizontal magnetic field vector, is one of SWPC’s primary prediction goals, so initial tests have been performed that compare modeled and observed \( \text{db/dt} \).

• Event: December 14, 2006 storm (Kp = 8 and Dst = -109.)
  Rationale: if testing one event, choose the one that is large, but not the strongest event. Strongest event, October 29, 2003, may yield atypical results. (Note, in these tests, the ranking includes all GEM events, but contingency tables are for 12/14/06 only).
Geomagnetic Stations Used for Test Runs

Table 2. The locations of the geomagnetic observatories used in the study.

<table>
<thead>
<tr>
<th>Station code</th>
<th>Geomagnetic latitude</th>
<th>Geomagnetic longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>YKC</td>
<td>68.9</td>
<td>299.4</td>
</tr>
<tr>
<td>MEA</td>
<td>61.6</td>
<td>306.2</td>
</tr>
<tr>
<td>NEW</td>
<td>54.9</td>
<td>304.7</td>
</tr>
<tr>
<td>FRN</td>
<td>43.5</td>
<td>305.3</td>
</tr>
<tr>
<td>IQA</td>
<td>74.0</td>
<td>5.2</td>
</tr>
<tr>
<td>PBQ</td>
<td>65.5</td>
<td>351.8</td>
</tr>
<tr>
<td>OTT</td>
<td>55.6</td>
<td>355.3</td>
</tr>
<tr>
<td>FRD</td>
<td>48.4</td>
<td>353.4</td>
</tr>
<tr>
<td>HRN</td>
<td>73.9</td>
<td>126.0</td>
</tr>
<tr>
<td>ABK</td>
<td>66.1</td>
<td>114.7</td>
</tr>
<tr>
<td>WNG</td>
<td>54.1</td>
<td>95.0</td>
</tr>
<tr>
<td>FUR</td>
<td>48.4</td>
<td>94.6</td>
</tr>
</tbody>
</table>

All stations listed are at high or mid-geomagnetic latitudes. High-latitude (pink) and mid-latitude (green) are the stations used in this test.
• Tests performed with all GEM Challenge model submissions.
• Compare modeled and observed $\frac{db}{dt}$ with 1-minute time steps.
• For visual inspection, generate model vs observations time series plots for each station and model.
• Compute **RMS differences and prediction efficiency** over various intervals (the entire event, several time segments?) and separately for mid- and high-latitude (Not done here, but may be useful for tracking model improvements.)
• Compute $\frac{db}{dt}$ from model and observation at 1 min cadence and use largest $\frac{db}{dt}$ in a 10-min, or similar, interval. Compute RMS differences and prediction efficiencies. This “windowing” may show improvements over 1-min comparisons and still have customer value. (Not done in this test.)
Threshold-based comparisons deemed possibly the most important type of metric to be used in the evaluation. Threshold-based comparisons between models and observations for \( \frac{db}{dt} \) (0.3 nT/s, 0.7 nT/s, 1.1 nT/s and 1.5 nT/s as used in Pulkkinen et al., 2010). Resulting contingency tables (45-min forecast window length used for tests) can be used to determine various skills such as probability of detection (POD), probability of false detection (POFD), Heideke skill score, critical success index…

Here we show POD- and POFD-based model rankings separately for mid and high latitude for different threshold levels.
Event no. 2, HIGH-LAT station YKC, model: 6_SWMF

Event no. 2, HIGH-LAT station PBQ, model: 6_SWMF

Event no. 2, HIGH-LAT station ABK, model: 6_SWMF

Blue – observations  Black -- model
Event no. 2, MID-LAT station NEW, model: 6_SWMF

Event no. 2, MID-LAT station OTT, model: 6_SWMF

Event no. 2, MID-LAT station WNG, model: 6_SWMF

Blue – observations  Black -- model
Contingency Table Example

EVENT NO. 2, DB/D T THRESHOLD 1.5 NT/S MODEL 6_SWMF

MID-LATITUDE

<table>
<thead>
<tr>
<th>FORECAST/OBSERVATIONS</th>
<th>YES</th>
<th>NO</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NO</td>
<td>5</td>
<td>136</td>
<td>141</td>
</tr>
<tr>
<td>TOTAL</td>
<td>5</td>
<td>136</td>
<td>141</td>
</tr>
</tbody>
</table>

HIGH-LATITUDE

<table>
<thead>
<tr>
<th>FORECAST/OBSERVATIONS</th>
<th>YES</th>
<th>NO</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES</td>
<td>17</td>
<td>8</td>
<td>25</td>
</tr>
<tr>
<td>NO</td>
<td>52</td>
<td>64</td>
<td>116</td>
</tr>
<tr>
<td>TOTAL</td>
<td>69</td>
<td>72</td>
<td>141</td>
</tr>
</tbody>
</table>

STANDARD CONTINGENCY TABLE FOR DICHOTOMOUS FORECASTS

<table>
<thead>
<tr>
<th>Event Forecast</th>
<th>Event Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>H (hit)</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>F (false alarm)</td>
</tr>
<tr>
<td>No</td>
<td>M (misses)</td>
</tr>
<tr>
<td>Yes</td>
<td>N (correct rejections)</td>
</tr>
<tr>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>
Geospace Model – Test POD Ranking
(example based on integration over all four event)

Probability of Detection
POD = H / (H+M)

Probability of False Detection
POFD = F / (N + F)
In some cases, models appear to be representing increases in observed magnetic field variations—perhaps better at mid-latitudes than high-latitudes on the night side (needs additional analysis).

Similar tests need to be performed for delta B.

For contingency tables:

- Thresholds need to be chosen carefully (Chris Balch studies will contribute to selection).
- Selection of appropriate skill measure. Several measures should be used in the final comparisons.

Demonstrates importance of regional forecasts:

- At same UT, observations (and model) results are very different at different local times—we can make the case for improvements with regional forecasts.

Need to determine spatial and temporal averaging/windowing achievable with models that still results in improved customer products.
Metrics
Regional K and db/dt
Regional dB/dt Prediction

Challenge

• How well can MHD models predict a regional (TBD) dB/dt (e.g. max disturbance, average disturbance, log-spectral distance) compared to the ground observed value over specified time interval (TBD)

• Currently Available: No product
Regional dB/dt Prediction

**Observations, Models**

<table>
<thead>
<tr>
<th>Observed dB/dt</th>
<th>MHD Model dB/dt</th>
</tr>
</thead>
<tbody>
<tr>
<td>At ground station (Regional)</td>
<td>At ground station (Regional)</td>
</tr>
</tbody>
</table>

**Compute skill (or other metric) for each model and compare**

Model dB/dt / Observed dB/dt
dB/dt Evaluation

Event x Model $y_i$
(Kp, Dst, LT of storm main phase...)

High Latitude
(repeat for mid-latitude)

Max 1-min db/dt
(10 minute window)

Contingency Table
(for different thresholds - (e.g. 1 nT/s, 1.5 nT/s...))

Skill metrics
(e.g. POD, Heideke, CSI, ETS, ...)

Ranking

• Decisions needed: select another event?, choose specific skill metric(s), include low latitude stations?, include more stations for better local time coverage?, select spatial and temporal averaging/windows, aggregate ranking over all events?, weighting high over mid lat?, weighting different skill metrics?,...
• 231 cases ≥ 100 nT/min
• 72 cases ≥ 200 nT/min
• 29 cases ≥ 300 nT/min
Probability for Max dB/dt to Exceed Thresholds vs Kp

<table>
<thead>
<tr>
<th>K</th>
<th>Prob &gt;50nT/m</th>
<th>Prob &gt;100nT/m</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>5</td>
<td>1.7</td>
<td>0.0</td>
</tr>
<tr>
<td>6</td>
<td>16.4</td>
<td>3.0</td>
</tr>
<tr>
<td>7</td>
<td>54.3</td>
<td>16.5</td>
</tr>
<tr>
<td>8</td>
<td>89.6</td>
<td>68.7</td>
</tr>
<tr>
<td>9</td>
<td>100.0</td>
<td>91.3</td>
</tr>
</tbody>
</table>

Courtesy C. Balch
Regional K Prediction

Challenge

• Can MHD models predict a regional (TBD) K that better represents a local geomagnetic disturbance than the currently available global Kp over specified time interval (TBD)?

• Currently Available: Wing Kp predicted from solar wind input at 15-min cadence and AF 3-hour near-real time Kp observed index
Regional K Prediction

Current Kp, Observations, Models

<table>
<thead>
<tr>
<th>Wing Kp, AF Kp (Global)</th>
<th>Observed $\Delta B$ At ground station (Regional)</th>
<th>MHD Model $\Delta B$ At ground station (Regional)</th>
</tr>
</thead>
</table>

Convert $\Delta B$’s to K values*

<table>
<thead>
<tr>
<th>Wing Kp, AF Kp</th>
<th>Ground Station Observed K</th>
<th>MHD model K</th>
</tr>
</thead>
</table>

Compute skill (or other metric) for each model

<table>
<thead>
<tr>
<th>Wing or AF Kp / Observed K</th>
<th>Model K / Observed K</th>
</tr>
</thead>
</table>

* Alternative: convert Costello and AF Kp’s to $\Delta B'$ at test station location; also need to consider valid latitude range (~48-62 deg) for K index
Regional delta B (K) Evaluation

- Event $x$ Model $y_i$ ($K_p$, Dst, LT of storm main phase…)
- Mid-latitude (repeat for high-latitude)
- $K_{obs}$, $K_{model}$, $K_{pWing}$ (3 hr interval, sliding 15 minute window)
- Contingency Tables (for different thresholds - (e.g. K>5, 6, 7 and 8)
- Skill metrics (e.g. POD, Heideke, CSI, ETS, …)
- Ranking and compare model to Wing Kp

- Decisions: select another event?, choose specific skill metric(s), include high-latitude stations?, include more stations for better local time coverage, select spatial and temporal averaging/windows, aggregate ranking over all events?, weighting high over mid lat, weighting different skill metrics, computing difference plots for scoring
Some Additional Questions and Issues

• Compare CCMC computation of \( \frac{db}{dt} \) and ground delta b’s with modeler techniques

• Do we need to remove model biases?

• High-resolution runs to assess improvements?

• What needs to be done to the model (or data stream) to work with imperfect data? Sensitivity of model to different input data streams and their errors?

• What is the model configuration operational goals? Are there things turned off in the models that could be turned on and how will this affect model performance?

• Use of models for future ionospheric or other products (e.g. auroral zone boundaries, magnetopause crossings, energy input to auroral zone…)

• NCO/EMC to SWPC file sizes

• Resources to run models at SWPC

• Rules of the road
• Model selection will be made by the SWPC Director

• Space Weather Prediction Testbed (SWPT) is responsible for making recommendations on candidate models

• SWPT will write a Recommendation Report based on internal/external team evaluation

• Modelers will have the opportunity to review and comment on the draft Recommendation Report prior to delivery to SWPC Director

• The final Recommendation Report will be made public
Possible Findings/Recommendations

• One (and only one) MHD model has sufficient value to justify transition and operation costs – Recommend transition

• Multiple MHD models have sufficient value – Recommend one model based on highest long-term value and lowest cost

• No MHD model has sufficient value, but near-term improvements could be made – Recommend SWPC support for additional development and testing

• One or both empirical models have sufficient value – Recommend either or both for transition

• No model has sufficient value – Recommend no SWPC action
Team Evaluation

• Internal participants: SWPT staff (2); Forecast Office (2); Development and Transition Section (2)

• External participants: Possibly including CCMC and others

• Evaluation Factors (roughly equal weight):
  - Strategic Importance
  - Operational Significance
  - Implementation Readiness
  - Cost to Operate, Maintain, and Improve
Strategic Importance

- Offers a critical new strategic opportunity
- Strengthens SWPC as quality focused and customer oriented
- Provides sufficient durability
- Minimizes duplication of effort
- Promotes SWPC/community relations
- Agreement on intellectual property rights
Operational Significance

- Prediction skill relative to Wing Kp
- High probability of detection and low false alarm rate
- Sizes of spatial and temporal windows with accurate results
- Additional validation needed to assess model potential
- Additional products that could be made available
- Possible improvement with higher resolution
Implementation Readiness

- Has reached a critical maturity level
- Facilitates efficient implementation on NOAA computers
  - Staff, funding, and time (internal and external) needed to implement and test
  - Documentation requirements
  - Training requirements
Cost to Operate, Maintain, and Improve

- Concept of Operations – How would the model be used?
- Level of effort required to run and monitor the model
- Effort required to implement on updated computer
- Short-term or long-term improvements envisioned
Next Steps
Geospace Model: Plans

- Plans (tentative schedule to be discussed with stakeholders)
  - 4/8 In-house discussion of plans
  - 4/8-25 Preliminary discussions and preparation for meeting with modelers
  - 4/25: All-day Geospace modeler meeting focused on evaluation metrics and selection process and initiate discussion to understand resource requirements
  - May-June: Refine and iterate metrics through test model runs
  - June 26 – July 1: GEM-CEDAR Workshop including Modeling Challenges and discussions with modelers on test runs
  - July - August 15: Model evaluation runs
  - August 15 – September 30 -- Interpretation and report preparation
  - Presentation and discussions at AGU on model run results
  - October – December: Model Selection at SWPC