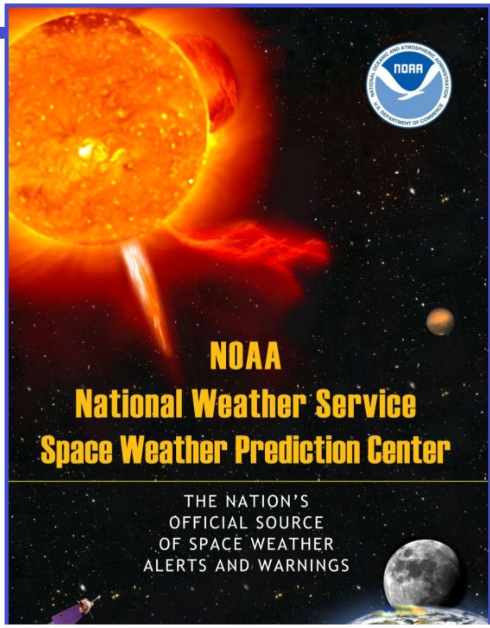
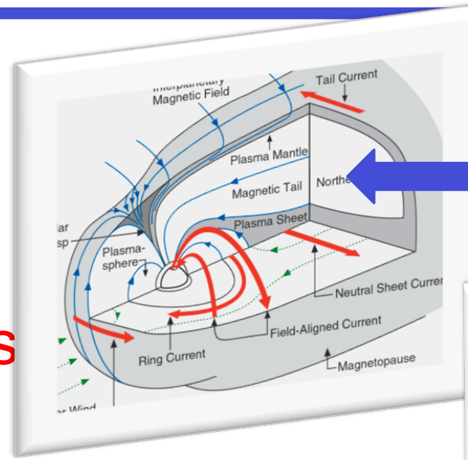
Geospace Model Workshop

April 25, 2011 Boulder, CO



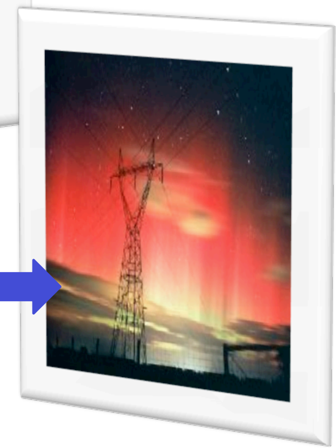
Today's Focus

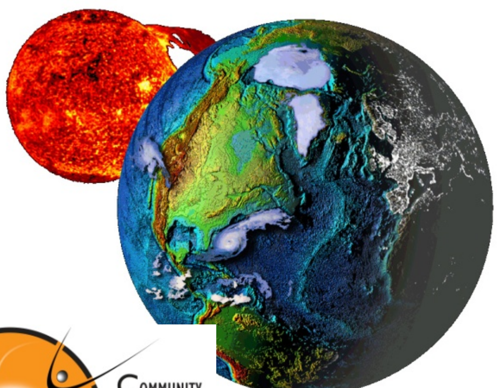
- Metrics
- Selection Considerations
- Plans



Geospace Models

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

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)

Geospace Model Workshop Agenda

April 25, 2011

Millennium Harvest House Hotel, Boulder, CO

0900 – 0915 **Introduction and Meeting Goals** – Singer and Kuznetsova

0915 – 0945 **Model Delivery to CCMC— Status, Validation, and Related Activities** – Kuznetsova

Delivered Model Configurations and discussion

0945 – 1000 **Open-GGCM** – Raeder

1000 – 1015 **CMIT** – Wiltberger

1015 – 1030 **SWMF** – Ridley

1030 – 1045 Coffee Break

1045 – 1100 **GUMICS** – Palmroth (presented by Singer and Pulkkinen)

1100 – 1115 **Weigel Empirical** – Weigel

1115 – 1130 **Weimer Empirical** – Weimer (presented by Singer)

1130 – 1200 **Metrics for CCMC Test Runs, Test Run results, and Discussion** (Singer and Pulkkinen)

1200 – 1330 Lunch

1330 – 1400 **Test Runs – continue discussion** (Singer and Pulkkinen)

1400 – 1500 **Proposed Metrics** – Singer
Metrics Discussion - All

1500 – 1600 **Selection Considerations** – Onsager
Selection Discussion – All

1600-1700 **Next Steps** – All

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).

• **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 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).

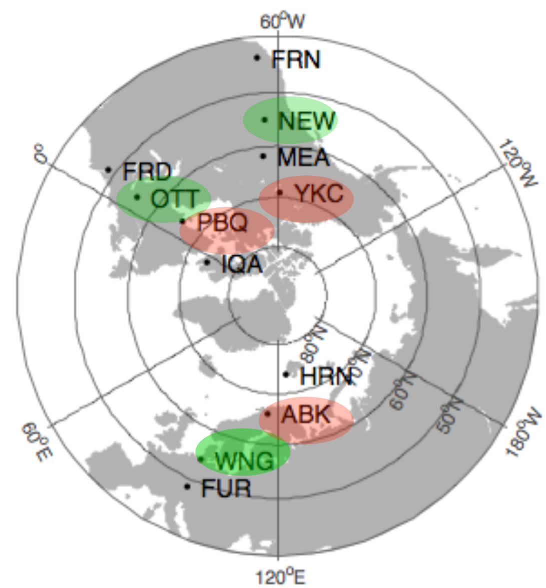
Geomagnetic Stations Used for Test Runs

PULKKINEN ET AL.: GEM CHALLENGE

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

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

PULKKINEN ET AL.: GEM CHALLENGE



. The locations and the station codes of the geomagnetic observatories used in

Geomagnetic dipole coordinates are used.

All stations listed are at high or mid- geomagnetic latitudes.

High-latitude (pink) and mid-latitude (green) are the stations used in this test.

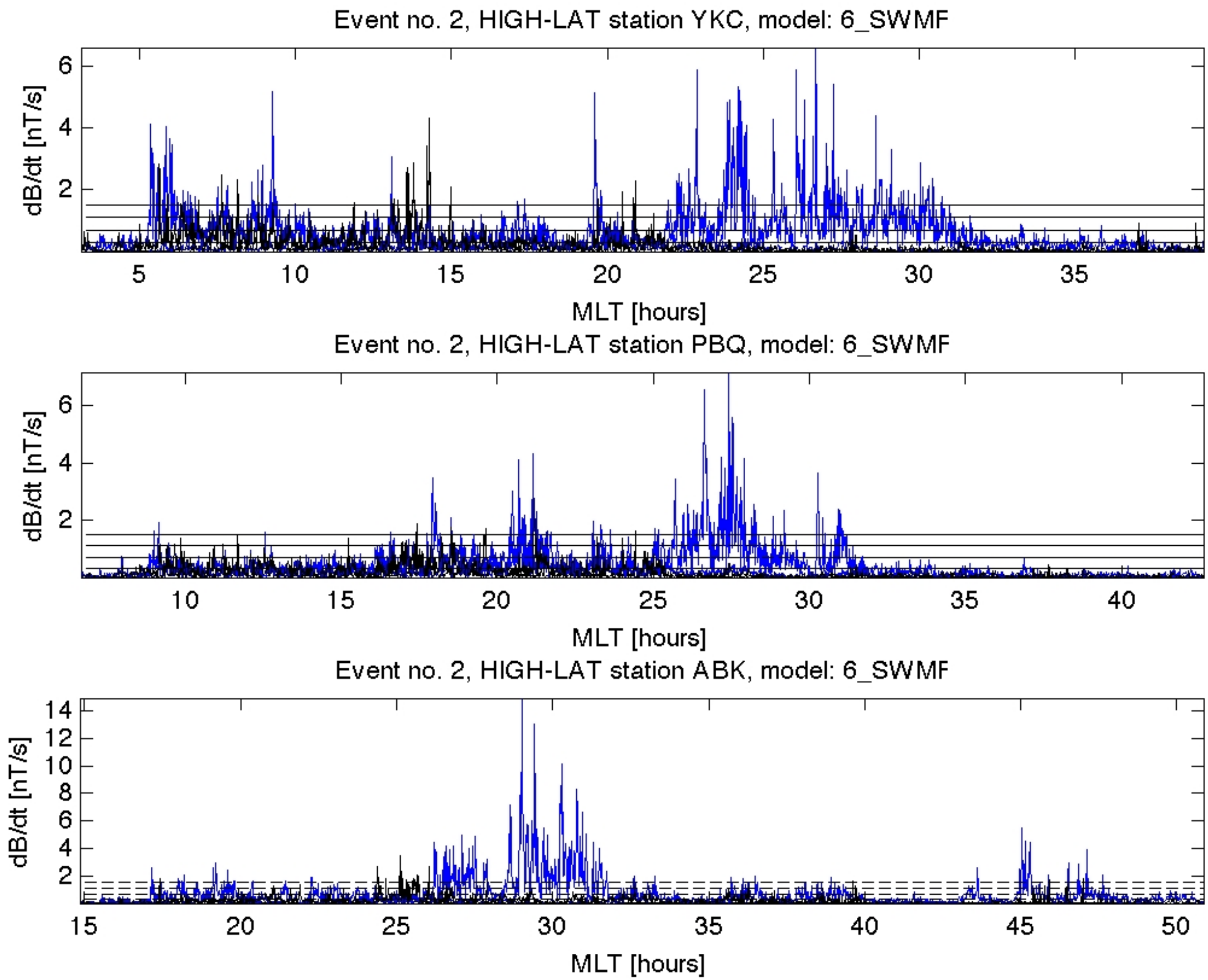
Geospace Model - Metric Tests

- Tests performed with **all GEM Challenge model submissions.**
- Compare modeled and observed **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 db/dt from model and observation at 1 min cadence and use largest 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.)

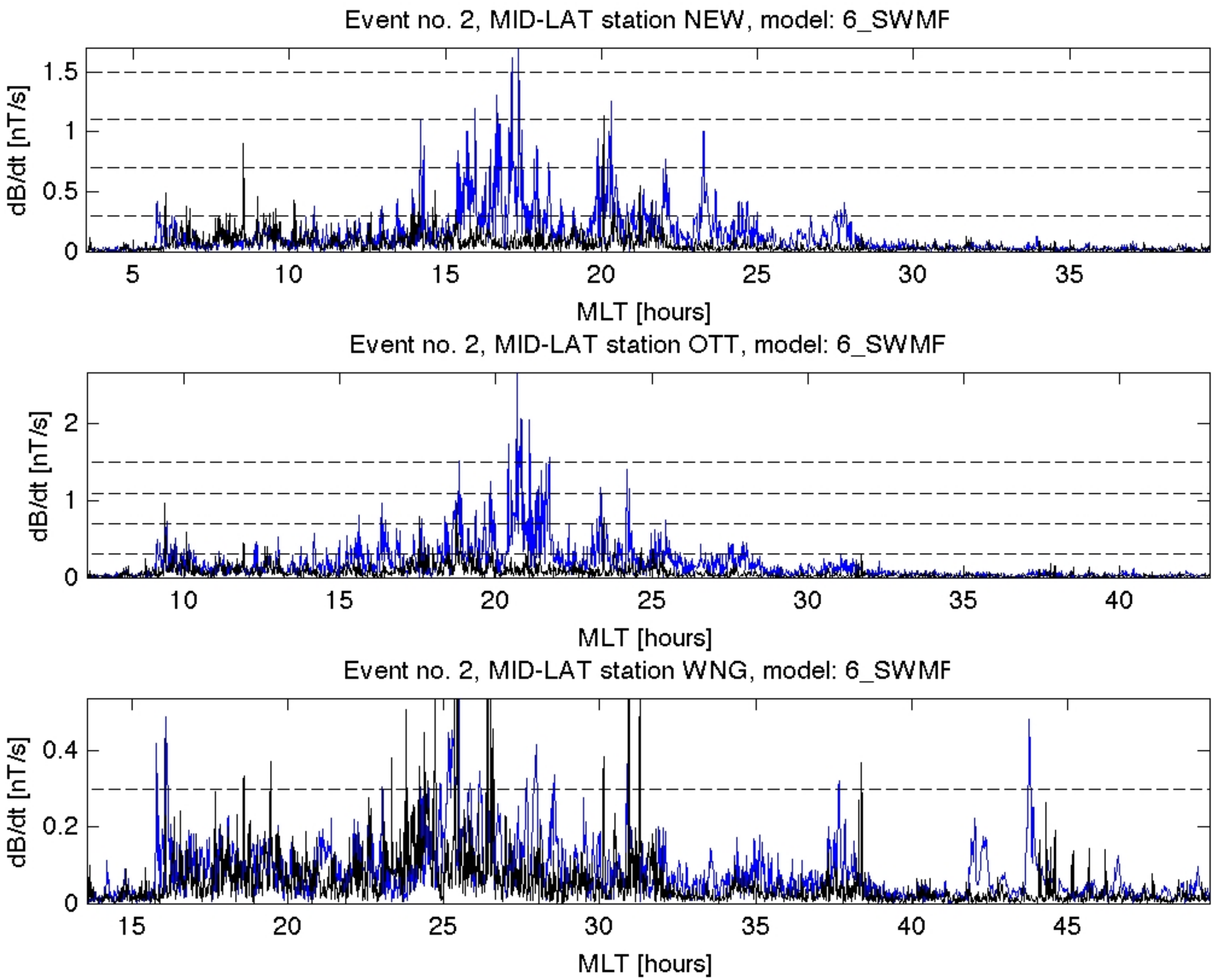
Geospace Model - Metric Tests

- **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 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.



Blue – observations Black -- model



Blue – observations Black -- model

Contingency Table Example

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

MID-LATITUDE

FORECAST/OBSERVATIONS	YES	NO	TOTAL
YES	0	0	0
NO	5	136	141
TOTAL	5	136	141

HIGH-LATITUDE

FORECAST/OBSERVATIONS	YES	NO	TOTAL
YES	17	8	25
NO	52	64	116
TOTAL	69	72	141

STANDARD CONTINGENCY TABLE FOR DICHOTOMOUS FORECASTS

Contingency Table		Event Observed	
		Yes	No
Event Forecast	Yes	H (hit)	F (false alarm)
	No	M (misses)	N (correct rejections)

Geospace Model – Test POD Ranking

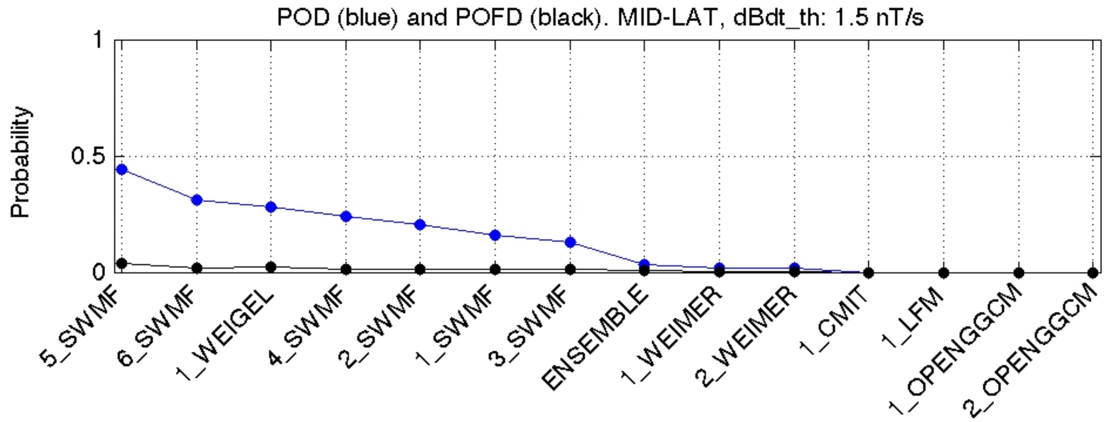
(example based on integration over all four event)

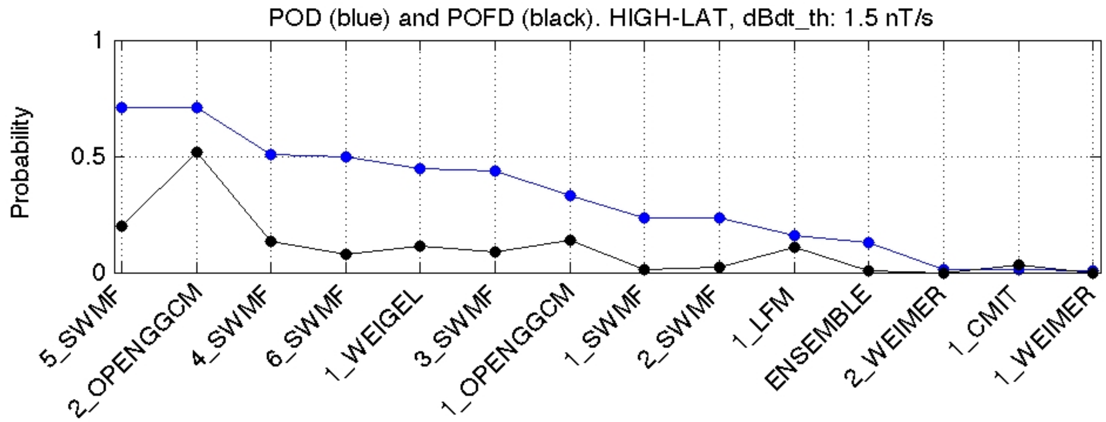
Probability of Detection

POD = H / (H+M)

Probability of False Detection

POFD = F / (N + F)





Geospace Model - Metric Tests Lessons Learned

- 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

Observed dB/dt At ground station (Regional)	MHD Model dB/dt At ground station (Regional)

↓

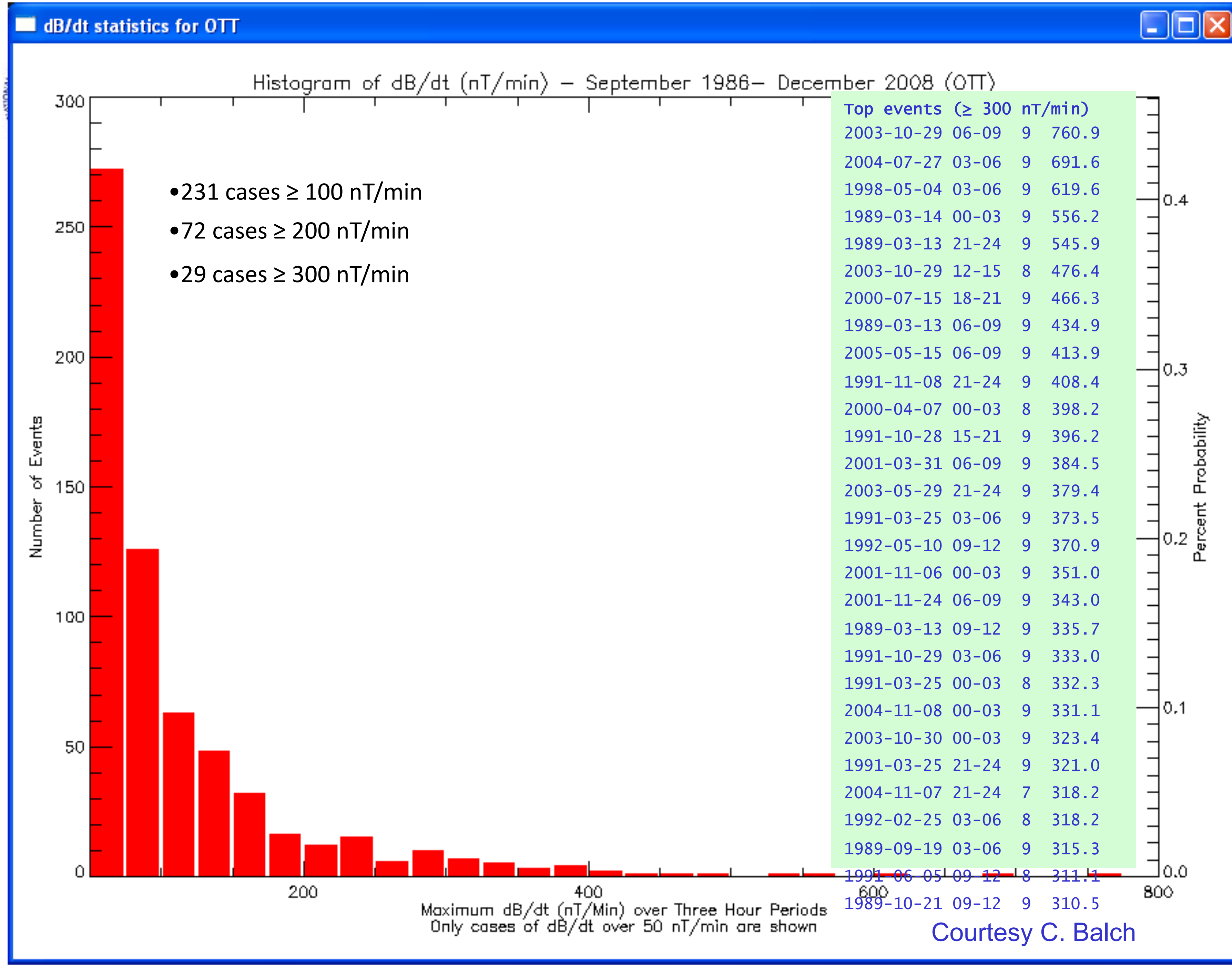
Compute skill (or other metric) for each model and compare

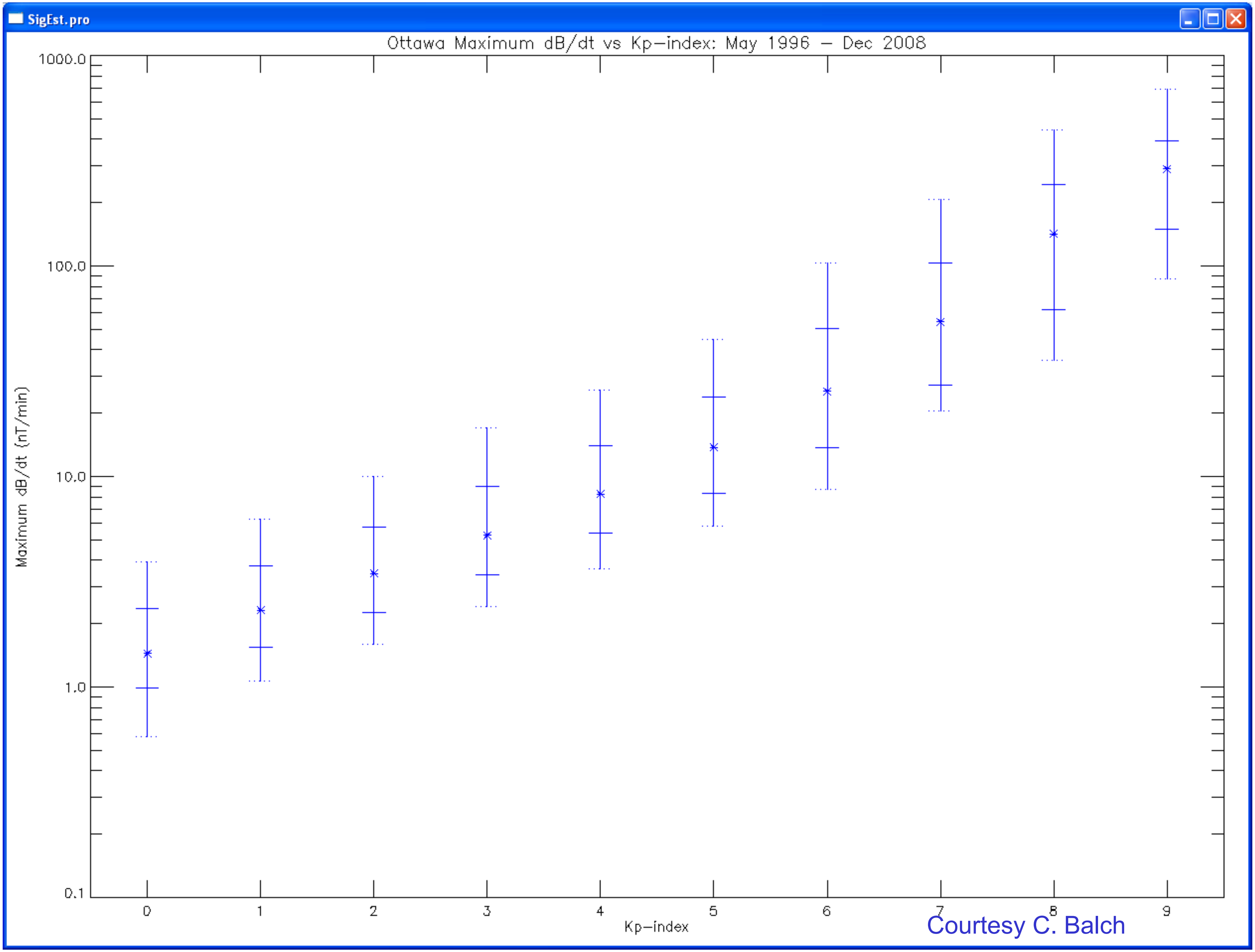
Model dB/dt / Observed dB/dt

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

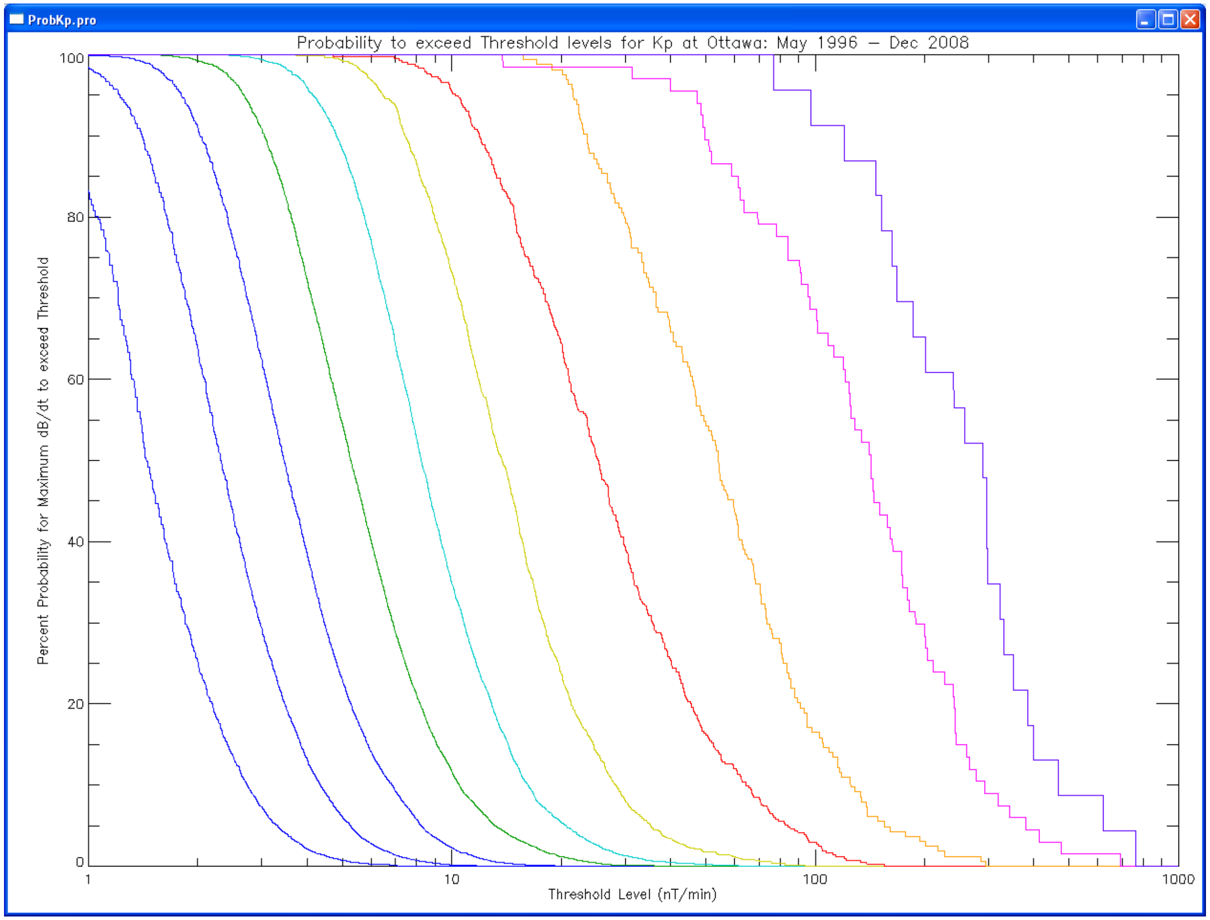
•Decisions needed: select another event?, choose specific skill metric(s), include low latitude stations?, include more stations for better local time coverage?, select spatial and temporal averaging/windows, aggregate ranking over all events?, weighting high over mid lat?, weighting different skill metrics?,…


dB/dt statistics for OTT
Histogram of dB/dt (nT/min) – September 1986– December 2008 (OTT)
•231 cases ≥ 100 nT/min
•72 cases ≥ 200 nT/min
•29 cases ≥ 300 nT/min
Number of Events
300
250
200
150
100
50
0
200
400
600
800
Maximum dB/dt (nT/Min) over Three Hour Periods
Only cases of dB/dt over 50 nT/min are shown
Percent Probability
0.4
0.3
0.2
0.1
0.0
Top events (≥ 300 nT/min)
2003-10-29 06-09 9 760.9
2004-07-27 03-06 9 691.6
1998-05-04 03-06 9 619.6
1989-03-14 00-03 9 556.2
1989-03-13 21-24 9 545.9
2003-10-29 12-15 8 476.4
2000-07-15 18-21 9 466.3
1989-03-13 06-09 9 434.9
2005-05-15 06-09 9 413.9
1991-11-08 21-24 9 408.4
2000-04-07 00-03 8 398.2
1991-10-28 15-21 9 396.2
2001-03-31 06-09 9 384.5
2003-05-29 21-24 9 379.4
1991-03-25 03-06 9 373.5
1992-05-10 09-12 9 370.9
2001-11-06 00-03 9 351.0
2001-11-24 06-09 9 343.0
1989-03-13 09-12 9 335.7
1991-10-29 03-06 9 333.0
1991-03-25 00-03 8 332.3
2004-11-08 00-03 9 331.1
2003-10-30 00-03 9 323.4
1991-03-25 21-24 9 321.0
2004-11-07 21-24 7 318.2
1992-02-25 03-06 8 318.2
1989-09-19 03-06 9 315.3
1991-06-05 09-12 8 311.1
1989-10-21 09-12 9 310.5
Courtesy C. Balch


SigEst.pro
Ottawa Maximum dB/dt vs Kp–index: May 1996 – Dec 2008
1000.0
100.0
10.0
1.0
0.1
Maximum dB/dt (nT/min)
0
1
2
3
4
5
6
7
8
9
Kp–index
Courtesy C. Balch

Probability for Max dB/dt to Exceed Thresholds vs Kp

K	Prob >50nT/m	Prob >100nT/m
0	0.0	0.0
1	0.0	0.0
2	0.0	0.0
3	0.0	0.0
4	0.7	0.0
5	1.7	0.0
6	16.4	3.0
7	54.3	16.5
8	89.6	68.7
9	100.0	91.3



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

Wing Kp, AF Kp (Global)	Observed ΔB At ground station (Regional)	MHD Model ΔB At ground station (Regional)

↓

Convert Δ B's to K values*

Wing Kp, AF Kp	Ground Station Observed K	MHD model K

↓

Compute skill (or other metric) for each model

Wing or AF Kp / Observed K	Model K / Observed K

* Alternative: convert Costello and AF Kp's to ΔB' 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 (Kp, Dst, LT of storm main phase…)

Mid-latitude (repeat for high-latitude)

K_{obs}, K_{model}, Kp_{Wing} (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 hight-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 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

Geospace Model Selection Process

- 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

End