Measuring the Performance of Scientific Models

K. A. Keller, M. Hesse, L. Rastaetter, M. M. Kuznetsova, T. Moretto
NASA Goddard Space Flight Center
Need for Metrics

- Create objective measure of current capabilities both for scientific and operational needs.
- Measure the improvement of model capabilities over time.
- Provide an objective comparison between models with comparable output.
  
  Metrics which lead to scores near unity now are useless!
Elements of a Metric

- An output parameter from a model.
  - An example is currents in the ionosphere can be used to calculate ground magnetic perturbations.

- A satellite or ground-based measurement that can be used for comparison.
  - An example is ground magnetometer data.

- A quantifiable norm that assesses the difference between the parameter from the model and the measurement.
Possible Metrics

- Ground magnetic perturbations using data from ground magnetometer chains.
- Particle fluxes at geosynchronous orbits using Los Alamos National Laboratory satellite data.
- Other metrics that may be suggested by the space weather operational or research community.
Community Coordinated Modeling Center (CCMC)

- Multi-agency partnership established to help bridge the gap between the space weather research community and operational agencies of National Oceanographic and Atmospheric Administration and the United States Department of Defense.
- Provides validation of models through both science-based testing and metrics evaluations by an independent evaluator.
- Serves the space weather research community by providing access to models through runs-on-request web site.
Ground Magnetic Perturbations

Data

- 10 stations in the Greenland chain using the H component of the data.

Models

- Weimer electric potential model (2 different versions).
- Weimer field-aligned current model (3 different versions).

Skill score

- An individual model is scored \( D_i = \sum |\Delta H_{\text{model}} - \Delta H_{\text{data}}|/npts \).
- A skill score is computed for each ground station by

\[
M_i = 1 - \frac{D_i}{D_s}
\]

where \( D_s \) is for the standard model. In this case, the standard model is \( \Delta H_{\text{standard}} \equiv 0 \).
Results for Weimer Models (averaged over 10 stations) for H component.

Score Averaged over 6 Days

Model and Version

- Weimer 2K Electric Potential Model
- Weimer Electric Potential Model Version 5
- Weimer Electric Potential Model Version 5 with MV delay
- Weimer FAC Model Version 1
- Weimer FAC Model Version 2
- Weimer FAC Model Version 5
- Weimer FAC Model Version 5 with MV delay
Parameter Tests

- Different time delays for the ACE data were used. The skill scores were not very sensitive to the time delays.

- Different Hall conductivities were used for the electric potential model. The skill scores were better for Hall conductivities of 5 and 7.5 mhos. For later versions, the scores are more sensitive to different conductivities.
Comparison of Model Results to Data

Black: Data from ground magnetometers
Orange: Model results from Weimer 2k Electric Potential Model
Blue: Model results from Weimer Electric Potential Model Version 5

Magnetometer data was provided by the Danish Meteorological Institute (Dr. Jurgen Watermann, Project Scientist)
Proton Fluxes

Data
- Proton fluxes from LANL geosynchronous satellites

Model
- Fok ring current model coupled to MHD models

Root Mean Square Error Skill Score
- Calculate root mean square error (RMSE)
  \[ \text{RMSE} = \sqrt{\frac{\sum (\text{predicted} - \text{observed})^2}{\text{npts}}} \]
- Calculate standard deviation of observations
  \[ \text{STD} = \sqrt{\frac{\sum (\text{observed} - \text{mean})^2}{\text{npts}}} \]
- RMSE skill score
  \[ \text{Skill score} = 1 - \frac{\text{RMSE}}{\text{STD}} \]

Cross Correlation
Sample of Ring Current Metric

<table>
<thead>
<tr>
<th>RMSE Skill</th>
<th>Cross Score</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.07</td>
<td>-.01</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Black is LANL data. Blue is the model results.

Geosynchronous proton flux data was provided by the Energetic Particle team at Los Alamos National Laboratory, Richard Belian (PI).
Summary

The ground magnetic perturbations is a first attempt at creation and application of a standard and repeatable metric.

Blind test (no fine tuning)!

Fine tuning of metrics is required in collaboration with the operational agencies and researchers.

First steps, more to come.