

Space Weather

RESEARCH ARTICLE

10.1029/2018SW002043

Special Section:

Space Weather Capabilities
Assessment

Key Points:

- We present an interactive online data-model analysis tool that is being constructed at the CCMC
- The online model-data analysis can span multiple events or observatories (of the same kind) to create aggregate skill scores
- Interactive plots allow users to inspect numerical values during their analysis

Correspondence to:

L. Rastaetter,
lutz.rastaetter@nasa.gov

Citation:

Rastätter, L., Wiegand, C., Mullinix, R. E., & MacNeice, P. J. (2019). Comprehensive Assessment of Models and Events Using Library Tools (CAMEL) framework: Time series comparisons. *Space Weather*, 17, 845–860. <https://doi.org/10.1029/2018SW002043>

Received 31 JUL 2018

Accepted 10 MAY 2019

Accepted article online 22 MAY 2019

Published online 17 JUN 2019

Published 2019. This article is a U.S. Government work and is in the public domain in the USA.

Comprehensive Assessment of Models and Events Using Library Tools (CAMEL) Framework: Time Series Comparisons

Lutz Rastätter¹ , Chiu P. Wiegand^{1,2}, Richard E. Mullinix^{1,3}, and Peter J. MacNeice¹ 

¹Community Coordinated Modeling Center, Code 674, NASA GSFC, Greenbelt, MD, USA, ²Code 550, NASA GSFC, Greenbelt, MD, USA, ³Code 587, NASA GSFC, Greenbelt, MD, USA

Abstract The Comprehensive Assessment of Models and Events using Library Tools (CAMEL) framework leverages existing Community Coordinated Modeling Center services: Run-on-Request postprocessing tools that generate model time series outputs and the new Community Coordinated Modeling Center Metadata Registry that describes simulation runs using Space Physics Archive Search and Extract metadata. The new CAMEL visualization tool compares the modeled time series with observational data and computes a suite of skill scores such as Prediction Efficiency, Root-Mean-Square Error, and Symmetric Signed Percentage Bias. Model-data pairs used for skill calculations are obtained considering a user-selected maximum difference between the time of observation and the nearest model output. The system renders available data for all locations and time periods selected using interactive visualizations that allow the user to zoom, pan, and pick data values along traces. Skill scores are reported for each selected event or aggregated over all events for all participating model runs. Separately, scores are reported for all locations (satellites or stations) and for each location individually. We are building on past experiences with model-data comparisons of magnetosphere and ionosphere model outputs from GEM2008, GEM-CEDAR Electrodynamics Thermosphere Ionosphere, and the SWPC Operational Space Weather Model challenges. The CAMEL visualization tool is demonstrated using three validation studies: (a) Wang-Sheeley-Arge heliosphere simulations compared against OMNI solar wind data, (b) ground magnetic perturbations from several magnetosphere and ionosphere electrodynamics models as observed by magnetometers, and (c) electron fluxes from several ring current simulations compared to Radiation Belt Storm Probes Helium Oxygen Proton Electron instrument measurements, integrated over different energy ranges.

1. Introduction

Model-data comparisons have been conducted at the Community Coordinated Modeling Center (CCMC) since 2008 when global magnetospheric models were run for selected geomagnetic storm events. The resulting outputs were compared against magnetic fields measured at GOES satellite locations as well as magnetometer locations on the ground (Pulkkinen et al., 2010; Rastätter et al., 2011). These studies explored model performance using a variety of skill scores, including Root-Mean-Square Error (RMSE), Prediction Efficiency (PE) for real data evaluations and threshold based (categorical or yes/no) evaluation metrics such as Probability of Detection (POD), Probability of False Detection (POFD), and Heidke Skill Score (HSS; Lopez et al., 2007; Pulkkinen et al., 2013).

Model validation efforts at CCMC also began in 2008 for the heliosphere where the impact of different magnetogram observations on Wang-Sheeley-Arge (WSA) heliosphere model (Arge & Pizzo, 2000) results near the Earth were studied (MacNeice, 2009a, 2009b). This study looked at hits and misses, timing errors, and minimum and maximum values of solar wind velocity jumps that identify crossings of the heliospheric current sheet.

Soon after these initial studies, observations in the ionosphere were collected to assess performance of ionosphere/thermosphere models in the CEDAR Electrodynamics Thermosphere Ionosphere (ETI) challenge (Shim et al., 2011, 2012). The GEM, CEDAR and SWPC study data and model outputs are available on the CCMC website utilizing a metrics visualization tool (called Metrics-Vis in this paper) that analyzes

data based on a selection of a single event (pre-defined time interval), observatory and physical parameter (described in Section 3.1).

The space weather forecasting activities at the CCMC since 2008 have resulted in extensive work in validating modeling of Coronal Mass Ejection (CME) arrival times at Earth using single and ensemble modeling at the CCMC (Mays et al., 2015; Taktakishvili et al., 2009) and the accumulation of community model results in a CME Scoreboard application (Riley et al., 2018; Wold et al., 2018; Verbeke et al., 2019)

Up until recently most of the model validations at the CCMC were available through Metrics-Vis. However, the tool was unable to aggregate multiple events or multiple observation locations (e.g., magnetometer locations) to calculate average skills or rank models by a single or a combination of multiple skills as was done in, e.g. (Pulkkinen et al., 2013; Rastätter et al., 2014a, 2013; Shim et al., 2011, 2012). The new Comprehensive Assessment of Models and Events Using Library Tools (CAMEL) toolkit expands on the capabilities of Metrics-Vis by presenting multiple observation locations and events in the same interface to allow for aggregation of skill scores across multiple events and stations or spacecraft locations that have been selected.

In recent years, data-model evaluations are being pursued in the fields of radiation belt, ring current, global magnetosphere ultralow frequency wave modeling and magnetopause location studies, extending the need for skill scores to support an ever-widening variety of data (Morley et al., 2018). In order to meet the needs of the community worldwide, the CCMC is developing the CAMEL framework (described in section 2) to support model-data comparisons more effectively and efficiently. The CAMEL visualization front end is being demonstrated using three use cases: Ground magnetic perturbation data from 13 magnetometers are simulated by three global magnetosphere and two statistical ionosphere electrodynamics models (Pulkkinen et al., 2013) for six space weather events. We show CAMEL's capability to collect and average skill scores across multiple events and stations in section 4.1. WSA (Arge & Pizzo, 2000; MacNeice, 2009a) heliosphere simulations with different prediction lead times compared to hourly OMNI2 (King & Papitashvili, 2005) solar wind data. We use CAMEL for a case where no events are defined and calculate skills for each month of a yearlong interval. Finally, RAM-SCB (Jordanova et al., 2010; Yu et al., 2011, 2019) and CIMI (Fok et al., 2014) ring current simulations are compared to Helium Oxygen Proton Electron (HOPE) mass spectrometer observations (Funsten et al., 2013; Spence et al., 2013) of electron fluxes on the Radiation Belt Storm Probes (RBSP or Van Allen Probes) for two geomagnetic storm events in section 4.3.

With the ring current data we employ the skill scores that equally penalize underestimations and overestimations of data by using the (Base-10) logarithm of the data to calculate scores like the Symmetric Signed Percentage Bias (SSPB; Morley et al., 2018).

Model-data validation efforts have been essential to terrestrial weather modeling where the Model Evaluation Tools (METs; ; Adriaansen et al., 2018; Halley Gotway et al., 2018) are being maintained in the Development Testbed Center (dtcenter.org) at the National Center of Atmospheric Research (NCAR). MET can calculate skill scores from time series, similar to efforts done in space weather research, and also performs pattern matching in 2-D and 3-D gridded data and model outputs (Davis et al., 2006a, 2006b; Prein et al., 2017; Wolff et al., 2014). CCMC has started a collaboration with NCAR to incorporate components of MET into the CAMEL framework in the future. Observations from two-dimensional images will be compared to gridded model output data in the future using tools in MET.

2. CAMEL Framework

The CAMEL framework is an integrated and flexible framework allowing users to seamlessly compare model outputs with similar observational data sets. The back end of the CAMEL framework takes advantage of existing services and will utilize newly-developed tools at the CCMC as they mature. The front end interactive visualization tool powered by the back end infrastructure allows users to plot model outputs and observation data sets together while providing options to calculate various skill scores from our library. In addition to driving the CAMEL interactive visualization tool presented in this paper, the CAMEL framework will provide an Application Programming Interface (API) allowing users to download all data sets available on the CAMEL back end database into their own environment for further analysis.

2.1. Leverage Existing CCMC Services

This is the list of existing CCMC services and systems that the CAMEL framework utilizes:

1. CCMC Metadata Registry (CMR; <https://kauai.ccmc.gsfc.nasa.gov/CMR>) contains metadata for CCMC models, model runs, and numerical and nonnumerical (e.g., image) outputs. The metadata data model is Space Physics Archive Search and Extract (SPASE) compliant (Roberts et al., 2018) allowing our system to easily connect to the SPASE registry for their registered observation data sets. All metadata for data sets stored in the CAMEL back end database are first registered in CMR. For model-data comparison and validation purposes, it is essential that the CAMEL framework has and can provide proper metadata regarding its observation data as well as model output to the users. We are continuing to expand the SPASE metadata model to support model chains and coupled frameworks that have become very common in the Space Physics community. We have regular weekly interactions with the SPASE group. Recently, our suggestions to add Developer, User, and HostContact as new roles were adopted by SPASE to better describe how the CCMC staff fit in support of the models we host. The SPASE hierarchy describes the model itself, its execution, and all input parameters and relevant run environment, and all output from running the model. Model type, region of space, and time interval are the primary parameters used to discover results, with use of more detailed metadata planned. Other suggested changes to SPASE are expected as we work to describe more complex simulations and their output.
2. Run-on-Request (RoR) system: The CCMC RoR system contains thousands of model runs and forms a wide archive of their outputs. One of the future goals of the CAMEL framework is to be able to utilize the existing RoR archive to validate models for the community. Currently, RoR runs are processed by hand and manually imported into the Metrics-Vis or CAMEL systems. CAMEL exists independently from RoR runs once their outputs have been ingested into CAMEL's database tables. In order to import RoR outputs automatically, a new next-generation RoR system (RoR-NextGen) is being designed and developed that will automatically register and access simulation metadata in the CMR. Therefore all existing and future model runs and output from the RoR system will contain proper metadata that is SPASE compliant. When RoR-NextGen has matured, the CAMEL framework will be able to connect to and search the archive of RoR runs at the CCMC allowing users to find and use all relevant output for their validation efforts.
3. Postprocessing tools: these include existing tools available to a user at CCMC such as the (online) visualization (CCMC-Vis) that performs extraction of time series data along satellite trajectories or at locations on the ground or in space. CalcDeltaB (Rastätter et al., 2014b) calculates magnetic perturbations at magnetometer stations from electric currents in the magnetosphere and ionosphere (available as a post-processing run request applied to any magnetosphere model simulation). Other tools at the CCMC include the internally developed and used Flexible Data Ingestion Tool, a generic data parser that reads time series data and stores them in a database table.

2.2. CAMEL Back End

The CAMEL back end infrastructure is designed and developed with flexibility in mind. The CAMEL database contains tables describing each validation study and captures information about the parameters that are relevant to a validation study (in terms of observation data and locations). Information regarding various validation studies are also captured in the database along with available predefined time intervals when applicable. All data sets are mapped to their appropriate SPASE descriptions identified by their SPASE ID in CMR. This allows the CAMEL back end to get detailed metadata regarding its data sets from CMR when necessary. Communication between the CAMEL back end and the interactive front end is implemented via a web service interface. The web service interface can also be used by other applications if users prefer to build their own visualization tools using the CAMEL data sets. It also provides an option for users to download and analyze data sets into their own favorite tools/framework.

One of the most powerful and crucial components of the CAMEL back end infrastructure is the library of skill scores, which may be applied to any set of observation-model-output pairs. Skill scores are single numerical values resulting from a summation of applicable samples within a chosen time range (event) or list of events. Samples are observation times where model output values were either matched in time or were close enough to be interpolated (given a tolerance entered by the user). Skill scores are tabulated for the events, simulation runs and spatial zones selected with a visualization. Currently, the CCMC is looking into incorporating the powerful METs from NCAR into the CAMEL framework. This would offer the space science community more calculations and formulas that the terrestrial weather community is already employing.

Before skill scores are calculated, data interpolation and filtering options are being applied. Interpolation options implemented are nearest neighbor and linear interpolation. Filters will apply smoothing and con-

Data and model comparisons

This is the web interface for the visualization of observational data and results of several model run results.

Please review the default selections below and make your changes.

To start the graphics program click the *Update Plot* button. The resulting image will be displayed at this location of the page.

Should the result be a black image, then the graphics program encountered a programming error. Please report the set of input parameters used.

[Go back to metrics challenge table](#)

Update Plot will update (generate) the plot with the chosen time and plot parameters below.

This will take some time (typically 10-30s) as data are read in and processed.

Start: Year: Month: Day: Hour: Minute: Second:
 to End: Year: Month: Day: Hour: Minute: Second:

Choose **Quantity** to be displayed:

Plot Options:

Image magnification

Line thickness (EPS image output only)

Character thickness (all annotations)

Thickness of Observation Data overplot (0: obs. data appears behind model results)

New: Logarithmic Y axis Yes (when log scale is selected, skills are calculated from the logarithm of observation and model values)

New: Apply lower limit to data Yes Minimum value:

New: Apply upper limit to data No Maximum value:

Vertical offset between data and model traces: (≠0: stack plot)

Lock plot range:

Min.: Max.:

Show scores

Spectral analysis: Window Length [min]: Window overlap [%]:

Select model settings

- red dark red blue green

Reset Form will reset changes to the defaults specified by the previous run of this script.

Update Plot will update (generate) the plot with the chosen time and plot parameters above.

Runs-on-Request: [Contact CCMC Staff](#)

Visualization: [Dr. Lutz Rastätter](#)

Figure 1. Community Coordinated Modeling Center (CCMC) time series comparison tool interface (“Metrics-Vis”) applied to electron fluxes observed at Radiation Belt Storm Probes spacecraft (https://ccmc.gsfc.nasa.gov/cgi-bin/run_metrics_vis.cgi?study=Plasma_Rad_Effects&event=1&metrics=2&obs=RBSPA-HOPE).

sider userselected minimum and maximum values for either observation or model results to be accepted in skill score calculations.

The back end (database server) runs at CCMC and is currently only available through the CAMEL front end run on a CCMC web server or a development server within the same private network. An API has not been formalized. The database server will be opened to outside networks once an API based on accepted web standards is defined and documented.

2.3. CAMEL Front End

The CAMEL front end focuses on the interactive visualization (described in the following section) that align observation data and model results at common times. Visualizations compare observations with data from one or more simulation runs or empirical model specifications and may include locations from multiple, similar instruments or satellites. Examples include magnetometers that can be divided into high-, middle-, and low-latitude locations or spacecraft of a series such as DMSP, GOES, or LANL that are located in different local time sectors or distances from Earth. Users can also select the skill scores that they would like to calculate based on their selected data sets. The front end of CAMEL also provides options for users to filter out data gaps and outliers if applicable. The CAMEL front end also offers a choice of interpolation method (nearest neighbor, linear interpolation) to use for model-data comparison as well as skill scores calculation. In this paper we present the advances in visualization and skill score calculations (with support from CMR in item 1 in section 2.1).

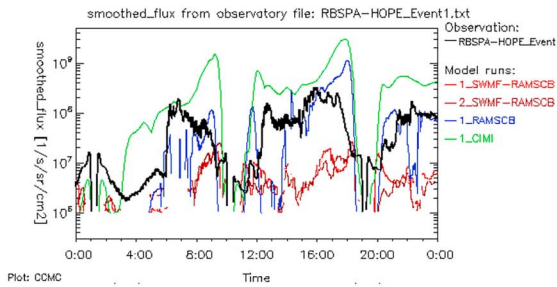


Figure 2. Metrics-Vis plot of electron fluxes at the RBSP-A spacecraft from HOPE instrument observations and SWMF-RAMSCB and stand-alone RAMSCB and CIMI model runs. The table of skill scores was generated with the plot by Metrics-Vis. RBSP = Radiation Belt Storm Probes; HOPE = Helium Oxygen Proton Electron.

EPS image (166.43 KB)

```
Variable: smoothed_flux Observatory file: RBSPA-HOPE_Event1.txt
Model Setting  PredEff  PredYield  Correlation  MinTimingError  MaxTimingError  RMS Error  SSPB  N_region  N_finite  LogSpecDist  nWin
1_SWMF-RAMSCB -2.303   0.5658    0.2922      0.8133          2.383          1.060     -162.8  4177     2619     NaN         0
2_SWMF-RAMSCB -2.946   0.5500    0.2120      9.983           6.295          1.118     -179.3  4177     2574     NaN         0
1_RAMSCB       -0.393   1.215     0.5303      11.24           2.112          0.6213    -3.602  4177     2333     NaN         0
1_CIMI         -1.133   1.385     0.6827      1.501           2.112          0.9208    90.40  4177     3823     NaN         0

PredEff      Prediction Efficiency metric
PredYield    ratio of the range of modeled values (max minus min) compared to the observation (max minus min)
Correlation  Cross-Correlation coefficient
MinTimingError absolute time difference between observation time of global minimum and time of modeled minimum (useful for events with single minimum value)
MaxTimingError absolute time difference between observation time of global maximum and time of modeled maximum (useful for events with single maximum value)
SSPB        Symmetric Signed Percentage Bias (Morley et al. (2018), Space Weather, 16, 69-88, doi:10.1002/2017SW001669)
N_region    number of samples in the selected time window
N_finite    number of points that were used for comparison (ie., those that were not NaN or infinite)
LogSpecDist Log-Spectral Distance metric
nWin        number of windows used for the spectral analysis (120-minute windows, offset by 30 minutes from the neighboring windows)
```

Figure 2. Metrics-Vis plot of electron fluxes at the RBSP-A spacecraft from HOPE instrument observations and SWMF-RAMSCB and stand-alone RAMSCB and CIMI model runs. The table of skill scores was generated with the plot by Metrics-Vis. RBSP = Radiation Belt Storm Probes; HOPE = Helium Oxygen Proton Electron.

3. Data-Model Comparisons and Event Assessments: Visualizations and Skill Scores

3.1. Old Online Time Series Visualization and Skill Score Calculation Tool

An online visualization and analysis tool has been available at the CCMC website via a *run_metrics_vis.cgi* interface since 2008, later referred in this paper as “Metrics-Vis.” The tool is available on the CCMC website (<https://ccmc.gsfc.nasa.gov>). On the “Metrics and Validation” page the user can select one of several challenges including the “GEM Challenge,” “CEDAR ETI Challenge,” or “GEM-CEDAR Challenge.” On each of those challenge pages, a “Time series plotting Tool” link directs the user to a table of visualization links utilizing the tool (e.g., “Time series plotting tool (ionosphere/thermosphere)” points to https://ccmc.gsfc.nasa.gov/challenges/GEM-CEDAR/plotting_tool_ion.php or “Time series plotting tool (magnetosphere)” points to https://ccmc.gsfc.nasa.gov/challenges/GEM-CEDAR/plotting_tool_mag.php from the GEM-CEDAR Challenge page). Each link on those tables provides a value to the tool for these four basic variables:

1. *Campaign*: the name of the validation study (e.g. GEM2008, CETI2010, SWPC2010, and Space Radiation and Plasma Effects);
2. *Observatory*: the name of the spacecraft, magnetometer, or other observatory/instrument used to obtain the observational data (e.g., GOES-10, RBSP-A, and YKC);
3. *Event*: the selected time interval from a list of available predefined events included in the study;
4. *Metrics*: the designation of a observational parameter and of applicable skill scores.

With these four variables the interface presents the user with a list of simulation runs that have been included as time series data into a file tree at the CCMC that can be visualized (see, e.g., Figures 1 and 2).

The user now can select the runs to include in the plot (traces shown in Figure 2, top) and table of skill scores (applicable scores displayed in Figure 2, bottom). Shell scripts that automate the plotting and skill score calculations across multiple events and locations may exist but they have been developed by the lead author for each study and are not available through CCMC to outside users. With CAMEL the aggregation of skill scores is now available within a single instance of the online interface where multiple events and observation locations can be selected before calculating scores. With the addition of observation data and model outputs in the future, CAMEL will apply to more Space Physics domains and provide more validation techniques in the future. Progress and results so far are described below.

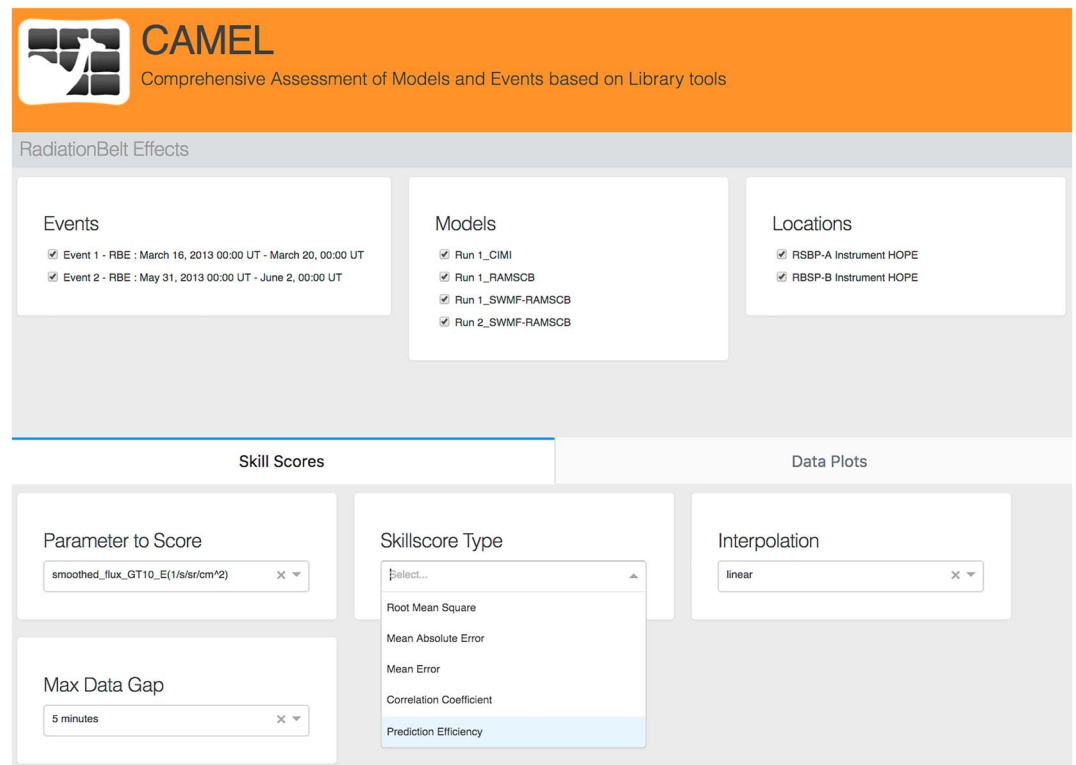


Figure 3. CAMEL Web Application User Interface for radiation belt effects study. With the domain and validation study preselected the interface offers choices of events, models, and observation locations in the first row. The second row of controls feature parameter, skill score (drop down activated to show choices), interpolation method, and time tolerance (maximum gap between nearest model output and observation time).

3.2. New CAMEL Interactive Visualization and Skill Scoring

The web application user interface for CAMEL, available through <https://ccmc.gsfc.nasa.gov/camel/> (Figure 3), provides applications for preselected values for Campaign and Metrics (the first and last of the four variables introduced in section 3.1). With the two variables selected, CAMEL retrieves the corresponding events, available simulation results, observatories and parameters names from the CAMEL database. The first row of controls in the interface (Figure 3) presents the user with a choice of events, simulation model runs, and observation locations. The second row then features choices of parameter, skill score, interpolation method, and maximum data gap.

The interpolation method and gap filtering based on a user-selected time interval are new features not available with the earlier “Metrics-Vis” plotting interface at CCMC. Skill scores are listed in one table for each location and across all selected locations or in another table for each time period (event) and across all selected events. Graphs are shown for each location (station, satellite) selected for one of the time periods or events. The data is graphed dynamically with panning, point picking (shown in Figure 4), and zooming capabilities.

As the tool chest of the CAMEL library grows, more granular functionality will be added to the user interface. These include rolling skill scores (based on interactive selection of start and stop times), scoring across multiple bases (observations from similar instruments, if available), smoothing options, and dynamic setting of threshold values to calculate skills based on hits and misses. Chaining of operations will allow for a flexible and powerful system that can be tailored to specific metrics. Each validation study uses a default analysis chain but CAMEL will allow for advanced tweaking of the skill score selection and of how some calculations are performed.

3.3. Skill Scores

The skill scores that apply to most real model output values y_i and observation data x_i at times i (model data interpolated to observation times) have been implemented in the old metrics validation system and



Figure 4. Comprehensive Assessment of Models and Events Using Library Tools plots of fluxes observed and modeled at the RBS-P-A and RBS-P-B locations. Point picking is showing a value along one of the model tracks. RBS-P = Radiation Belt Storm Probes; HOPE = Helium Oxygen Proton Electron.

are repeated in the new interactive visualization (with $\langle x \rangle = \frac{1}{N} \sum_{i=1}^N x$ being the arithmetic mean and $\sigma(x) = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \langle x \rangle)^2}$ the standard deviation of a times series x):

- Root-Mean-Square Error (RMSE): $RMSE = \sqrt{\langle (y_i - x_i)^2 \rangle}$,
- Correlation Coefficient (CC): $CC = \frac{\langle (x_i - \langle x \rangle)(y_i - \langle y \rangle) \rangle}{\sigma(x)\sigma(y)}$,
- Prediction Efficiency (PE): $PE = 1 - \frac{\langle (x_i - y_i)^2 \rangle}{\sigma^2(x)}$.
- Mean Absolute Error (MAE): $MAE = \langle |y_i - x_i| \rangle$
- Median Absolute Error (MdAE): $MdAE = M(|y - x|)$ with $M(e)$ denoting the median value of a distribution e .

Skills to be implemented include the Mean Absolute Percentage Error (MAPE): $MAPE = 100/N \sum_{i=1}^N |(y_i - x_i)/x_i|$ (see, e.g., Morley et al., 2018).

3.4. Examples of SPASE Model Run Description

Having proper metadata for all data sets is an important first step for model-data comparison. Metadata contain searchable details of the model version and configuration used that may include descriptions of component models in a coupled simulation, input parameters and numerical settings used in the simulation. Within the CAMEL framework, all metadata stored is SPASE compliant. An example of such is provided in XML format in Figure 5 for the WSA model version 2.2.

In Figure 6 below, it shows the metadata for the predicted plus 1-day solar wind parameters generated by the near-real-time WSA version 2.2 model run. From the metadata of the output, one can see that it is generated by the parent resource which in this case is the WSA v.2.2 model. With the linkage information in the metadata model, it is easy for anyone to see what the output is and which model and model run generated such output. This information is crucial for others to verify and be able to reproduce the same results.

4. Results

4.1. Ground Magnetic Perturbations at 13 Magnetometer Stations

Three global magnetosphere models and two statistical models were run to specify magnetic perturbations on the ground that are caused by changes in the magnetosphere and ionosphere of the Earth (Pulkkinen et al., 2011, 2013; Rastätter et al., 2014a). Solar wind data for six selected events lasting from 1 to 2 days each were selected and observations at 13 magnetometer stations in “chains” (listed north to south) on the

```

<SimulationModel>
  <ResourceID>spase://CCMC/SimulationModel/WSA/v.2.2</ResourceID>
  <ResourceHeader>
    <ResourceName>WSA</ResourceName>
    <AlternateName>The Wang-Sheeley-Arge model </AlternateName>
    <ReleaseDate>2017-01-01T00:00:00Z</ReleaseDate>
  <Description>
    The Wang-Sheeley-Arge (WSA) model has two components: the WSA Potential Field + Current Sheet (WSA PF+CS) Model and the WSA Inner Heliosphere (WSA-IH) Model. The WSA-PF+CS model is the inner coronal component of the complete WSA model. These combine a Potential Source Surface Model with the 'Schatten' current sheet model, to produce a model of the global coronal magnetic field between the solar surface and a bounding spherical surface, typically set at 5 solar radii. The magnetic field is assumed to be radial at this outer surface. At this outer surface the model computes the solar wind speed using an empirical relationship based on the divergence of the magnetic field and the proximity of the selected open field line to the nearest coronal hole boundary, to determine the local solar wind speed. The surface magnetic field is determined from synoptic magnetogram data. Currently we download magnetograms from both Kitt Peak and Mount Wilson.
  </Description>
  <Acknowledgement>Air Force Office of Scientific Research</Acknowledgement>
  <Contact>
    <PersonID>spase://CCMC/Person/Peter.MacNeice</PersonID>
    <Role>ModelHostContact</Role>
  </Contact>
  <Contact>
    <PersonID>spase://CCMC/Person/Charles.Nickolos.Arge</PersonID>
    <Role>ModelDeveloper</Role>
  </Contact>
</ResourceHeader>
<ModelVersion>
  <version>v.2.2</version>
  <changeLog/>
</ModelVersion>
<caveats/>
<SimulationType>
  <SimulationType>Empirical</SimulationType>
</SimulationType>
<CodeLanguage/>
<TemporalDependence>false</TemporalDependence>
<SimulatedRegion>
  <Name>Heliosphere.Inner</Name>
</SimulatedRegion>
  <Name>Sun.Corona</Name>
</SimulatedRegion>
</SimulationModel>

```

Figure 5. Wang-Sheeley-Arge (WSA) v.2.2 model metadata in Space Physics Archive Search and Extract format.

American West Coast (Yellowknife [YKC], Meanook [MEA], Newport [NEW], and Fresno [FRN]), East Coast (Iqaluit [IQA], Poste-de-la-Baleine [PBQ] or Sanikiluaq [SNK; for later events], Ottawa [OTT], and Fredericksburg [FRD]), and Europe (Hornsund [HRN], Abisko [ABK], Wingst [WNG], and Furstenfeldbruck [FUR]) were compared with results obtained from the models. Magnetic perturbation data (ΔB_{North} , ΔB_{North} , ΔB_{Down}) are compared to the observations for stations poleward of the auroral zone (YKC, IQA, and HRN), near the auroral zone (high latitude; MEA, PBQ/SNK, and ABK), subauroral latitudes (midlatitudes, $\sim 55^\circ$; NEW, OTT, and WNG), and at low latitudes ($< 50^\circ$; FRN, FRD, and FUR).

The CAMEL tool allows the user to select subsets of stations and events to perform ranking of models and calculate averages of skill scores over the set of stations per event and over all events. Skill scores selected from a range of options are tabulated, and rankings visualized. For reference each model-data comparison plot for each event contains the traces (observation and model traces) for each magnetometer in a subpanel in the plot for each event. In the following we present results from CAMEL applications for selected validation campaigns. The first is the GEM2008 Ground magnetic perturbation study. Figure 7 shows a sample analysis on ΔB_{North} performed with CAMEL showing the skill ranking of the different models for all stations. To refine the analysis the user can select one of the available quantities (ΔB_{North} , ΔB_{East} , and ΔB_{Down}) and any subset of the 8 events and the 13 magnetometer stations. Figure 8 shows scores for auroral-latitude stations (ABK, PBQ, and MEA) for Event 1. Skill scores are calculated for each selected model run, event, and station, then tabulated for all selected stations and each station individually (across all selected events) and for all events and each event (for the set of selected stations). This allows the user to evaluate how each model performs for the different events and also by selecting different subsets of the stations, where models perform better or worse spatially.

The graphing capability is shown in Figure 9 for station IQA during Event 1. The CAMEL tool allows the user to show model and data comparison plots for one event at a time at multiple station locations (in separate plot panels, not shown).

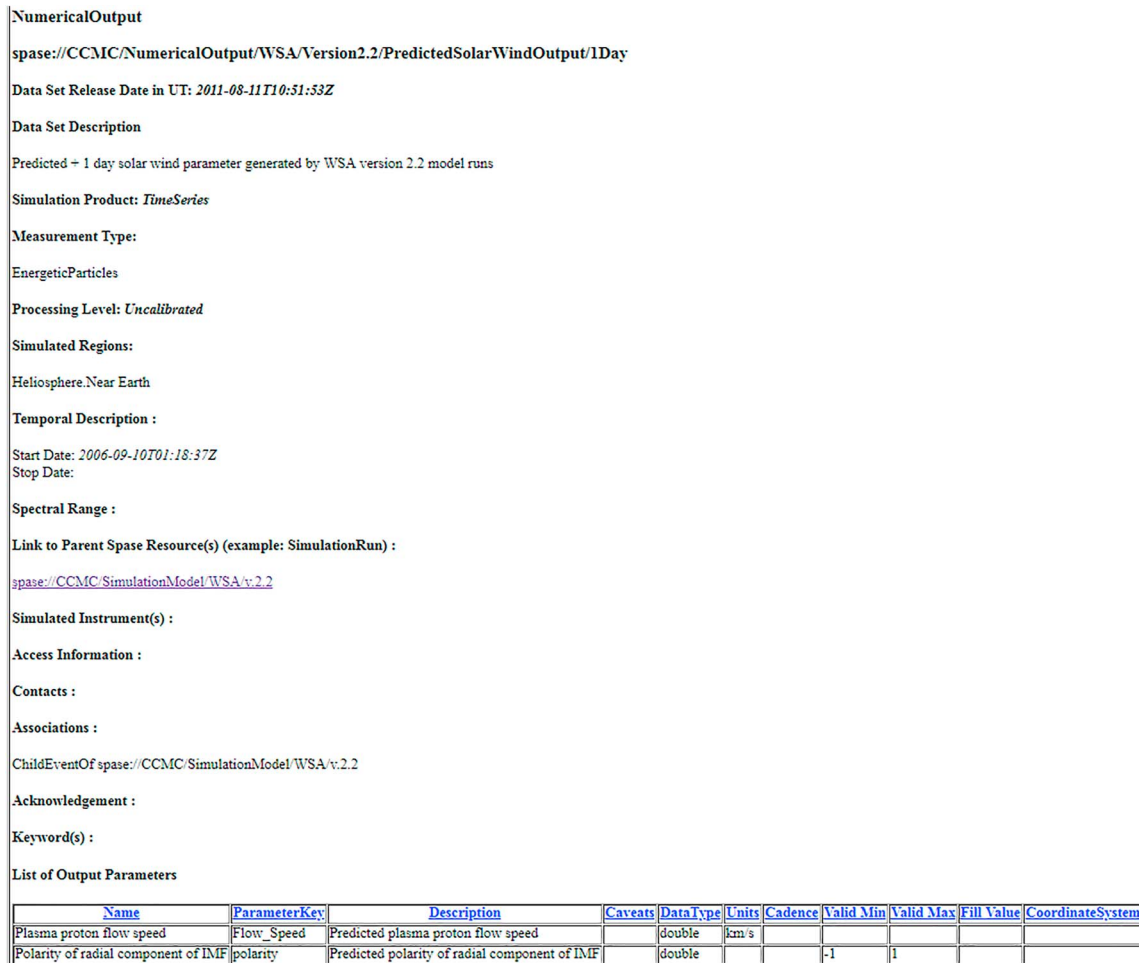


Figure 6. Wang-Sheely-Argé (WSA) v.2.2 model output for the predicted +1-day solar wind parameters.

In the near future we plan to implement bar graphs to visualize the skill scores listed in the tables. Later in the development of the tool we envision the option to evaluate models using two skill scores. Visualization would render the models' performance in a two-dimensional (Taylor-type) plots.

Once we import time derivatives of the horizontal component of the magnetic perturbations ($d\Delta B_H/dt$; Pulkkinen et al., 2013), we will add a user interface to select threshold values with a list of the values that were used in the published study to generate a contingency table of Hits, Misses, False Predictions, and Correct Negatives. We will calculate skills such as POD, POFD, or HSS.

4.2. WSA Model and In Situ Observation of Density and Velocity at L1 in the Heliosphere

The WSA (Argé & Pizzo, 2000) near-real-time “simulation” is a continuous series of coronal and heliospheric solutions that are constructed by the of the WSA model based on magnetograms of the Sun that are updated every 6 hours. Each WSA calculation in that series provides forecasts with lead times between 1 and 7 days (in daily increments) of solar wind speed and density at the first Lagrange point near Earth (about 240 Earth Radii or about 1.6×10^6 km closer to the Sun from Earth on the Sun-Earth line). Metadata stored for such a run can be found in the CMR. Solar wind monitoring satellites such as IMP8, WIND (Lepping et al., 1995; Lin et al., 1995), ACE, and DSCOVR (McComas et al., 1998; Smith et al., 1998) reside in circular orbits around L1 and their data are aggregated into the OMNI2 database. Model outputs are compared to OMNI2 hourly averages of velocity and density.

The CAMEL visualization tool interpolates model outputs to the times of the observation data points. The user can select between linear interpolation and nearest neighbor interpolation. Gaps in model data are eliminated by selecting any maximum time between model outputs that should be used to perform interpolations. To find matching model outputs, the nearest neighbor is sought among the model outputs for each

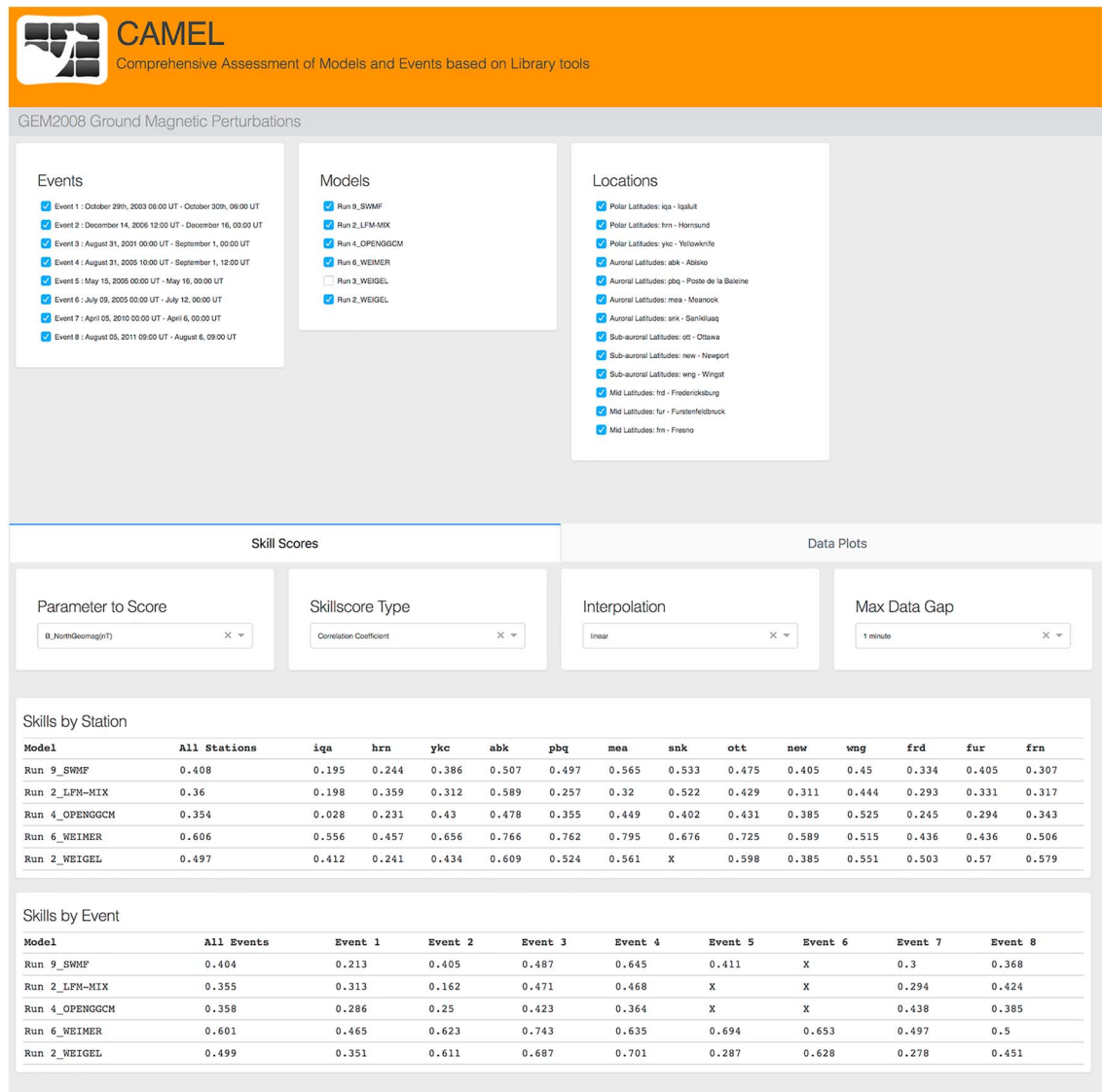


Figure 7. Cross-Correlation scores for ΔB_{North} at all magnetometers for all events for the five simulation runs 9_SWMF, 4_OpenGGCM, 2_LFM-MIX, 2_WEIGEL, and 6_WEIMER. Skill scores are tabulated for each model by station (averaged over all selected events) and by event (averaged over all selected stations).

observation data time. A matching pair is created only if the difference between the observation time and the nearest model output time is at or below the user-selected time threshold. Model outputs are then interpolated using the selected method (e.g., nearest neighbor interpolation, linear interpolation) to create the modeled value in the matched pair. All created pairs are then used in skill score calculations. The percentage of available samples is the ratio between the number of matched pairs and the number of observation times for a given time interval (event). In the near future we plan to add the capability of listing the percentage of matched samples that were used in each skill score calculation. Time intervals indicating Carrington Rotations can be presented for the users to use as events.

Figure 10 shows the 1-day and 4-day predictions from WSA compared to the 1-hr OMNI data. This month's skill scores indicate average agreement between observation and models (see Table 1).

4.3. Radiation Belt Dynamics

As part of an ongoing assessment of model skill specifying electron flux in the inner magnetosphere, several ring current and radiation belt models are being run for selected time periods (including 17 March 2013 and 31 May to 1 June 2013). Modeled energy distributions of electron fluxes are interpolated to energies

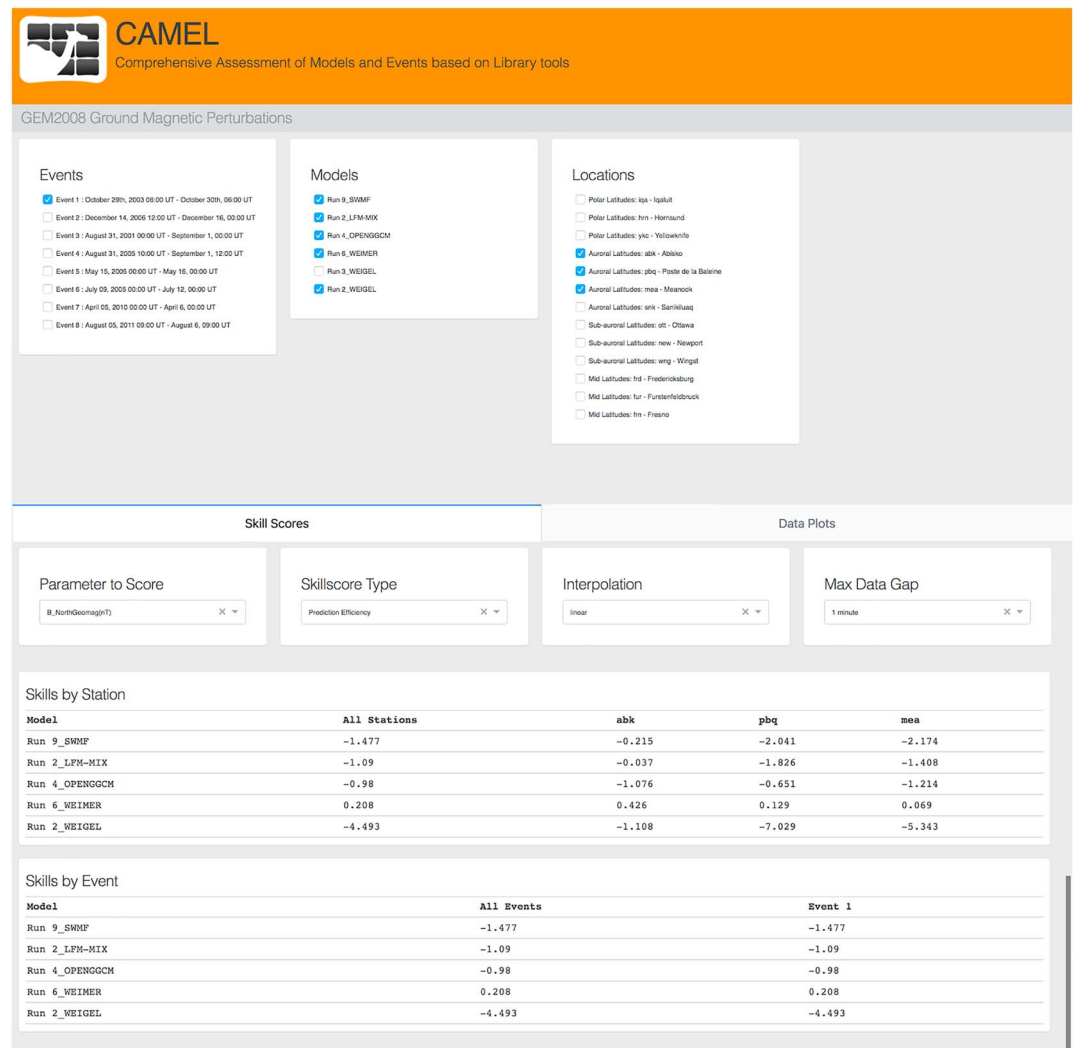


Figure 8. Prediction Efficiency scores for ΔB_{North} at auroral-zone magnetometers (ABK, PBQ, and MEA) for the Event 1 for five simulation runs 9_SWMF, 4_OpenGGCM, 2_LFM-MIX, 2_WEIGEL, and 6_WEIMER. This mode of operation can generate scores by region for multiple events similar to Figure 7.

measured by the HOPE instrument on the two RBSP (or Van Allen Probes) satellites (RBSP-A and RBSP-B) that orbit between 2- and 7-RE distance around the Earth. HOPE energies range from a few electron volts to 40 keV and electron fluxes at energies above 10 keV are correlated with surface charging events in Low-Earth Orbiting satellites (Yu et al., 2019). The online visualization and analysis tool on the CCMC website plots electron fluxes on a logarithmic scale and applies differences of logarithms of flux values to calculate skill scores. This ensures that skills are calculated based on ratio of observation and model values instead of linear differences, which are meaningless in an environment that is characterized by differences spanning orders of magnitude (up to and beyond 10^6 ; ; Friedel et al., 2002; Selesnick & Blake, 1997). Morley et al. (2018) introduced the Median Symmetric Accuracy and the SSPB, improvements over the Mean Absolute Percentage Error scores that are based on $\log(\text{flux})$ values.

The skill scores in Table 2 were generated with the old *run_metrics_vis.cgi* interface on the CCMC website (with study = Plasma_Rad_Effects obs=RBSPA-HOPE or RBSPB-HOPE, event = 1 or 2, and metrics=2 to compare integrated fluxes over 10 keV [metrics = 1 compares integrated flux over all HOPE energies], using https://ccmc.gsfc.nasa.gov/cgi-bin/run_metrics_vis.cgi?study=Plasma_Rad_Effects&event=2&metrics=2&obs=RBSPB-HOPE). PE and CC scores are available in CAMEL but SSPB scores have not yet been added.

Figure 11 shows the electron fluxes over all energies observed during the day of 17 March 2013 at RBSP-A (purple) and four different model runs (1_SWMF-RAMSCB [green], 2_SWMF-RAMSCB [same model with

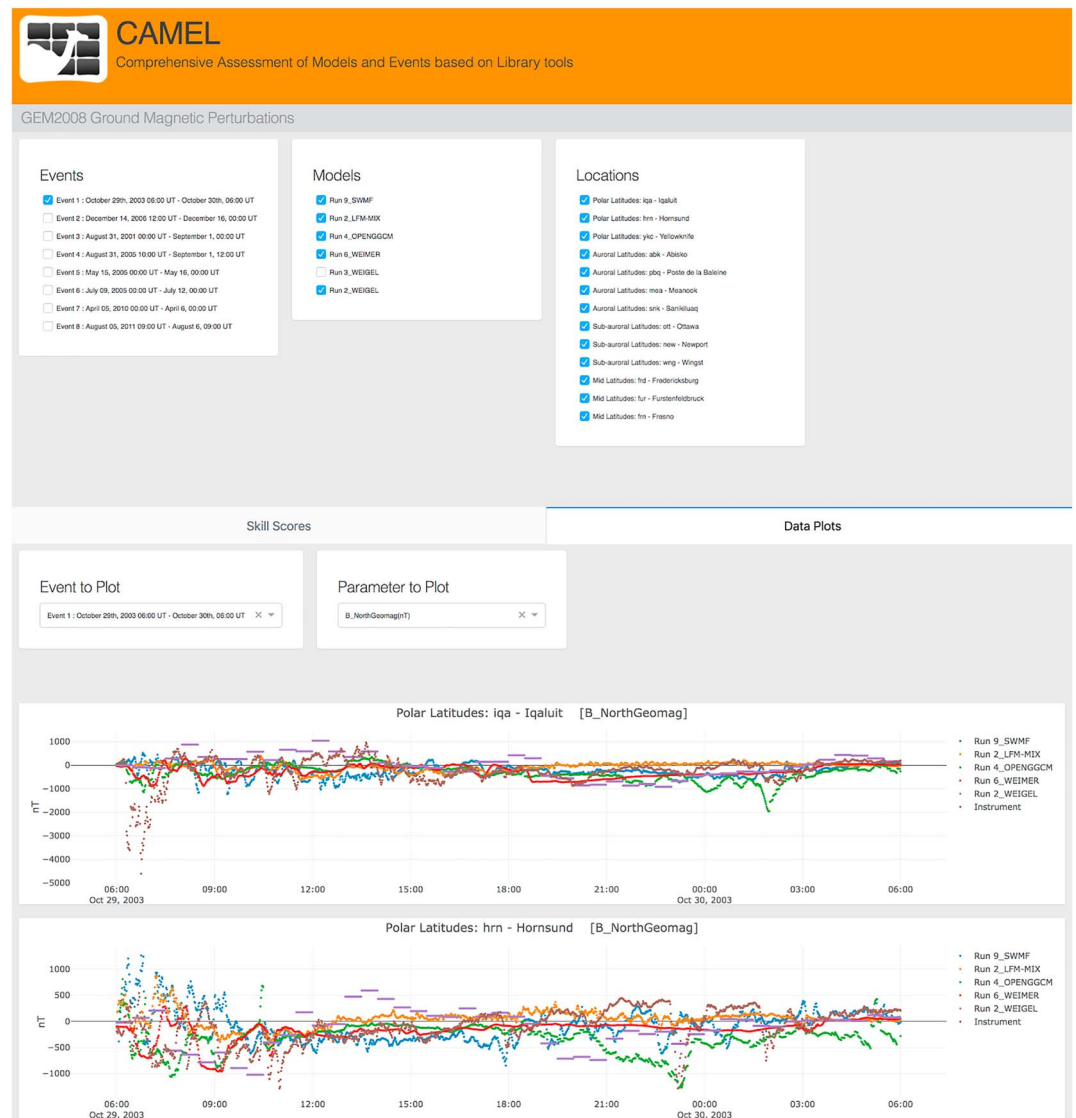


Figure 9. Time series plots of ΔB_{North} at Iqaluit (IQA) for Event 1 for the five simulation runs. A separate interactive graph is created for each station selected (two graphs are shown). The user can query numerical values by mouse-over and select a portion of the graph for detailed display (not shown).

slight adjustments, red], 1_CIMI [blue], and 1_RAMSCB [stand-alone RAMSCB, orange]). We implemented in Metrics-Vis the skill score $SSPB = 100\text{sgn}(M(\ln(y/x)))(\exp(|M(\ln(y/x))|) - 1)$ with $M(z)$ denoting the median value of the argument z (array of individual time step values) as described in Morley et al. (2018). This skill score penalizes overestimation by a given factor by the same weight as underestimation. It is insensitive to outlying values as it uses the median value while providing an intuitive percentage value. $SSPB$ is well-suited for observations where values vary over a large dynamic range as found in the Earth's radiation belt. In CAMEL we plan to offer $SSPB$ and the related score $\zeta = 100(\exp(M(|\ln(y/x)|)))$, both introduced by Morley et al. (2018), for studies of radiation belt and ring current fluxes.

5. Summary and Outlook

We demonstrated essential new capabilities of the CAMEL framework to support comprehensive model-data comparison studies. CAMEL is capable of calculating skill scores across multiple events (pre-defined time periods) and across multiple station or spacecraft locations to generate aggregate skill scores at each location for all selected events and at all locations for each event. Plots of data and model results can be created for multiple locations, if applicable, for one of the selected time periods. In addition, CAMEL

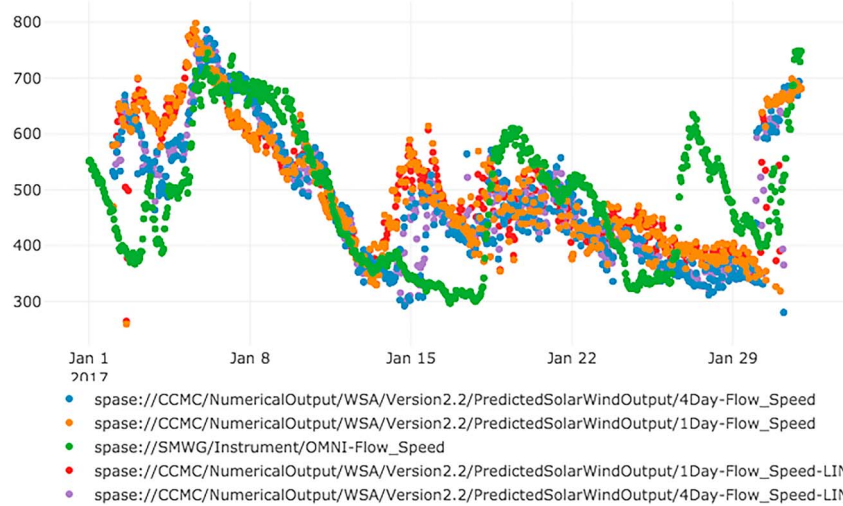


Figure 10. Wang-Sheeley-Arge 1-day (orange: original data, red: linear interpolated) and 4-day predictions (blue, purple) and ONMI solar wind speed (green) for January 2017. CCMC = Community Coordinated Modeling Center.

allows for a selection of interpolation methods and how long of a gap between observation data and nearest model output in time to accept when performing interpolations. All capabilities enable the user to perform validation studies beyond the reach of the previously existing Metrics-Vis interface available at the CCMC website that was restricted to a single event and a single observation location.

The CAMEL framework to compare observation data with model results is still undergoing rapid development. We plan to have a single CAMEL interface that can feature all available validation studies by physical domain and type of parameter to cover observations and model outputs in the Solar Corona, Heliosphere, Magnetosphere, Ionosphere/Thermosphere and on the Earth's surface. We are working to add more skill scores and interpolation methods and to provide a user interface to apply minimum and maximum values when making plots and calculating skills. We envision the CAMEL framework to extend into multi-dimensional data-model comparisons by employing pattern and feature extraction tools. The capabilities of MET at NCAR that is used for National Oceanic and Atmospheric Administration weather modeling are being investigated for possible inclusion. Later, pattern recognition software within MET that weather forecasters use to track two-dimensional features such as clouds or storm fronts in time will be used in model-model comparisons or instances where imaging is available on a regular basis such as in CME modeling in the heliosphere where comparisons with Coronagraph images or heliospheric Jmaps (space-time plots tracking features moving along a certain line in the plane of view) can be performed. Pattern recognition and tracking will also be useful in comparison of ionosphere and auroral observations and modeling.

The interactive visualization and skill scoring system will include categorical comparisons such as were done for the ground magnetic perturbation study where levels of dB/dt on a 1-min time scale from first-principles models of the global magnetosphere-ionosphere system and some statistical models of magnetic perturbation were compared to observations at selected magnetometers on the ground using predefined thresholds of dB/dt , a table of Hits, Misses, False Positives, and Correct No-forecasts were assembled and skills such as

Table 1
Monthly Skill Scores for 2016 (June–December) and 2017 (January–June) From the WSA Model's Solar Wind Speed Predictions With 1 and 4 Days of Lead Time

Run	Skill	July	Aug.	Sep.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May	June ^a	Year
1 day ^b	CC	0.293	0.260	0.691	0.700	0.622	0.490	0.511	0.707	0.406	0.282	0.737	0.557	0.519
4 day	CC	0.117	0.125	0.744	0.778	0.523	0.482	0.646	0.740	0.427	0.059	0.648	0.642	0.498
1 day	PE	-0.159	-0.211	0.457	0.428	0.230	0.171	0.077	0.417	0.025	-0.193	0.539	0.117	0.191
4 day	PE	0.511	-0.418	0.507	0.575	-0.0903	0.123	0.319	0.445	0.116	-0.805	0.393	0.321	0.133

Note. CC = Correlation Coefficient; PE = Prediction Efficiency; WSA = Wang-Sheeley-Arge.

^aComparisons with only 30% model coverage. ^bRuns labeled by number of days of lead time rather than the Space Physics Archive Search and Extract ID.

Table 2

Skill Scores for the Two Events (16 March 2013 00:00 to 20 March 2013 00:00 UT and 31 May 2013 00:00 UT to 2 June 2013 00:00 UT) in the Modeling of Integrated Fluxes Measured by RBSP (VAP) HOPE Instrument for Energies Above 10 keV

Run ^a	SPCR	Skill					
		CC		PE		SSPB	
		Event 1 ^b	Event 2 ^c	Event 1	Event 2	Event 1	Event 2
1_CIMI	RBSP-A	0.683	0.252	-1.133	-13.896	90.4	38.7
1_RAMSCB	RBSP-A	0.530		-0.393		-3.6	
1_SWMF-RAMSCB	RBSP-A	0.292	0.395	-2.303	-1.038	-162.8	-121.6
2_SWMF-RAMSCB	RBSP-A	0.212		-2.946		-179.3	
1_CIMI	RBSP-B	0.699		-2.201		145.7	
1_RAMSCB	RBSP-B	0.285		-11.242		-156.9	
1_SWMF-RAMSCB	RBSP-B	0.453	0.415	-2.470	-0.685	-137.5	-32.6
2_SWMF-RAMSCB	RBSP-B	0.326		-3.110		-156.8	

Note. Skills were calculated using logarithmic fluxes when model and observation values were $\geq 10^6$ /s/sr/cm². CC = Correlation Coefficient; PE = Prediction Efficiency; RBSP = Radiation Belt Storm Probes; VAP = Van Allen Probes; HOPE = Helium Oxygen Proton Electron.

^aRuns as labeled in the old visualization. ^b17 March 2013 00:00 to 18 March 2013 00:00 UT. ^c31 May 2013 00:00 to 2 June 2013 00:00 UT.

POD, of POFD, and HSS were calculated to rank the models. Together with event definitions the tool will allow the user to control threshold levels and subdivision of time intervals into sampling windows as was done for the previous work.

The coordination with VxOs via SPASE is out of scope of the CAMEL project. In the future a VxO interface to CCMC simulation runs will be mediated by the future RoR-NextGen system using SPASE descriptions of model runs and output parameters generated by each of the models. CAMEL's support of SPASE descriptors will eventually enable a direct pairing of observations with model results for automatic analyses.

RSBP-A Instrument HOPE [Smoothed integrated particle flux for all energies]

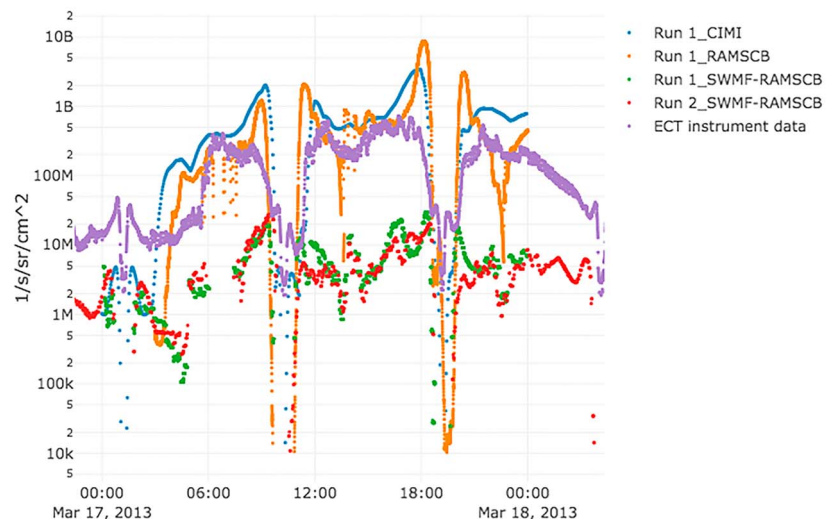


Figure 11. RBSP-A HOPE integrated electron flux over all energies (purple) and model results (other colors) for the day of 17 March 2013 (part of Event 1). RBSP = Radiation Belt Storm Probes; HOPE = Helium Oxygen Proton Electron.

Acknowledgments

This work is performed as part of research at the NASA Goddard Space Flight Center's Community Coordinated Modeling Center and is funded in part by the National Science Foundation (NSF) Space Weather Program. Data used in the model-data comparisons are the 1-hr OMNI2 data set for the solar wind (<https://omniweb.gsfc.nasa.gov/>) and ECT-HOPE data in the radiation belt from the Van Allen Probes (RBSP, https://rbsp-ect.newmexicoconsortium.org/data_pub/).

References

- Adriaansen, D., Win-Gildenmeister, M., Frimel, J., Prestopnik, J., Gotway, J. H., Jensen, T., et al. (2018). The METplus version 2.0 user's guide. Developmental Testbed Center.
- Arge, C. N., & Pizzo, V. J. (2000). Improvement in the prediction of solar wind conditions using near-real time solar magnetic field updates. *Journal of Geophysical Research*, *105*(A5), 10,465–10,479. <https://doi.org/10.1029/1999JA000262>
- Davis, C. A., Brown, B., & Bullock, R. (2006a). Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Monthly Weather Review*, *134*, 1772–1784. <https://doi.org/10.1175/MWR3145.1>
- Davis, C. A., Brown, B., & Bullock, R. (2006b). Object-based verification of precipitation forecasts. Part II: Application to convective rain systems. *Monthly Weather Review*, *134*, 1785–1795. <https://doi.org/10.1175/MWR3146.1>
- Fok, M.-C., Buzulukova, N. Y., Chen, S.-H., Glocer, A., Nagai, T., Valek, P., & Perez, J. D. (2014). The comprehensive inner magnetosphere-ionosphere model. *Journal of Geophysical Research: Space Physics*, *119*, 7522–7540. <https://doi.org/10.1002/2014JA020239>
- Friedel, R. H., Reeves, G. D., & Obara, T. (2002). Relativistic electron dynamics in the inner magnetosphere—A review. *Journal of Atmospheric and Solar-Terrestrial Physics*, *64*, 265–282. [https://doi.org/10.1016/S1364-6826\(01\)00088-8](https://doi.org/10.1016/S1364-6826(01)00088-8)
- Funsten, H. O., Skoug, R. M., Guthrie, A. A., MacDonald, E. A., MacDonald, J. R., Harper, R. W., et al. (2013). Helium, Oxygen, Proton, and Electron (HOPE) mass spectrometer for the Radiation Belt Storm Probes mission. *Space Science Reviews*, *179*, 423. <https://doi.org/10.1007/s11214-013-9968-7>
- Halley Gotway, J., Newman, K., Jensen, T., Brown, B., Bullock, R., & Fowler, T. (2018). The Model Evaluation Tools v8.0 (METv8.0) user's guide. Developmental Testbed Center.
- Jordanova, V. K., Zaharia, S., & Welling, D. T. (2010). Comparative study of ring current development using empirical, dipolar, and self-consistent magnetic field simulations. *Journal of Geophysical Research*, *115*, A00J11. <https://doi.org/10.1029/2010JA015671>
- King, J. H., & Papitashvili, N. (2005). Solar wind spatial scales in and comparisons of hourly wind and ACE plasma and magnetic field. *Journal of Geophysical Research*, *110*, A02104. <https://doi.org/10.1029/2004JA010649>
- Lepping, R. P., Acuña, M. H., Burlaga, L. F., Farrell, W. M., Slavin, J. A., Schatten, K. H., et al. (1995). The WIND magnetic field investigation. *Space Science Review*, *71*, 207–229. <https://doi.org/10.1007/BF00751330>
- Lin, R. P., Anderson, K. A., Ashford, S., Carlson, C., Curtis, D., Ergun, R., et al. (1995). A three-dimensional plasma and energetic particle investigation for the WIND spacecraft. *Space Science Review*, *71*, 125–153. <https://doi.org/10.1007/BF00751328>
- Lopez, R. E., Hernandez, S., Wiltberger, M., Huang, C. L., Kepko, E. L., Spence, H., et al. (2007). Predicting magnetopause crossings at geosynchronous orbit during the Halloween storms. *Space Weather*, *5*, 01005. <https://doi.org/10.1029/2006SW000222>
- MacNeice, P. (2009a). Validation of community models: Identifying events in space weather model timelines. *Space Weather*, *7*, S06004. <https://doi.org/10.1029/2009SW000463>
- MacNeice, P. (2009b). Validation of community models: 2. Development of a baseline using the Wang-Sheeley-Arge model. *Space Weather*, *7*, S12002. <https://doi.org/10.1029/2009SW000489>
- Mays, M. L., Taktakishvili, A., Pulkkinen, A., MacNeice, P. J., Rastätter, L., Odstroil, D., et al. (2015). Ensemble modeling of CMEs using the WSA-ENLIL+Cone model. *Solar Physics*, *290*(6), 1775–1814. <https://doi.org/10.1007/s11207-015-0692-1>
- McComas, D. J., Bame, S. J., Barker, P., Feldman, W. C., Phillips, J. L., Riley, P., & Griffie, J. W. (1998). Solar wind electron proton alpha monitor (SWEPAM) for the advanced composition explorer. *Space Science Review*, *86*, 563. <https://doi.org/10.1023/A:1005040232597>
- Morley, S. K., Brito, T. V., & Welling, D. T. (2018). Measures of model performance based on the log accuracy ratio. *Space Weather*, *16*, 69–88. <https://doi.org/10.1002/2017SW001669>
- Prein, A. F., Liu, C., Ikeda, K., Bullock, R., Rasmussen, R. M., Holland, G. J., & Clark, M. (2017). Simulating north american mesoscale convective systems with a convection permitting climate model. *Climate Dynamics*. <https://doi.org/10.1007/s00382-017-3993-2>
- Pulkkinen, A., Kuznetsova, M. M., Rastätter, L., Singer, H., Balch, C., Weimer, D., et al. (2013). Community-wide validation of ground magnetic field perturbation predictions of geospace models to support model transition to operations. *Space Weather*, *11*, 369–385. <https://doi.org/10.1002/2013SW000990>
- Pulkkinen, A., Kuznetsova, M. M., Ridley, A., Raeder, J., Vapirev, A., Weimer, D., et al. (2011). Geospace environment modeling 2008–2009 challenge: Ground magnetic field perturbations. *Space Weather*, *9*, S02004. <https://doi.org/10.1029/2010SW000600>
- Pulkkinen, A., Rastätter, L., Kuznetsova, M. M., Hesse, M., Ridley, A., Raeder, J., et al. (2010). Systematic evaluation of ground and geostationary magnetic field predictions generated by global magnetohydrodynamic models. *Journal of Geophysical Research*, *115*, A03206. <https://doi.org/10.1029/2009JA014537>
- Rastätter, L., Kuznetsova, M. M., Glocer, A., Welling, D., Meng, X., Raeder, J., et al. (2013). Geospace Environment Modeling 2008–2009 Challenge: D_{st} index. *Space Weather*, *11*, 20036. <https://doi.org/10.1002/swe.20036>
- Rastätter, L., Kuznetsova, M. M., Tóth, G., & Pulkkinen, A. (2014a). CalcDeltaB: An efficient postprocessing tool to calculate ground-level magnetic perturbations from global magnetosphere simulations. *Space Weather*, *12*, 553–565. <https://doi.org/10.1002/2014SW001083>
- Rastätter, L., Kuznetsova, M. M., Vapirev, A., Ridley, A., Wiltberger, M., Pulkkinen, A., et al. (2011). Geospace environment modeling 2008–2009 challenge: Geosynchronous magnetic field. *Space Weather*, *9*, S04005. <https://doi.org/10.1029/2010SW000617>
- Rastätter, L., Tóth, G., Kuznetsova, M. M., & Pulkkinen, A. A. (2014b). CalcDeltaB: An efficient postprocessing tool to calculate ground-level magnetic perturbations from global magnetosphere simulations. *Space Weather*, *11*, 553–565. <https://doi.org/10.1002/2014SW001083>
- Riley, P., Mays, M. L., Andries, J., Amerstorfer, T., Biesecker, D., Delouille, V., et al. (2018). Forecasting the arrival time of coronal mass ejections: Analysis of the CCMC CME scoreboard. *Space Weather*, *16*, 1245–1260. <https://doi.org/10.1029/2018SW001962>
- Roberts, D. A., Thieman, J., Génot, V., King, T., Gangloff, M., Perry, C., et al. (2018). The SPASE data model: A metadata standard for registering, finding, accessing and using heliophysics data obtained from observations and modeling. *Space Weather*, *16*, 1899–1911. <https://doi.org/10.1029/2018SW002038>
- Selesnick, R. S., & Blake, J. B. (1997). Dynamics of the outer radiation belt. *Geophysical Research Letters*, *24*, 1347–1350. <https://doi.org/10.1029/97GL51409>
- Shim, J. S., Kuznetsova, M., Rastätter, L., Bilitza, D., Butala, M., Codrescu, M., et al. (2012). CEDAR Electrodynamics Thermosphere Ionosphere (ETI) challenge for systematic assessment of ionosphere/thermosphere models: Electron density, neutral density, NmF2, and hmF2 using space based observations. *Space Weather*, *10*, S10004. <https://doi.org/10.1029/2012SW000851>
- Shim, J., Kuznetsova, M., Rastätter, L., Hesse, M., Bilitza, D., Butala, M., et al. (2011). CEDAR electrodynamics thermosphere ionosphere (ETI) challenge for systematic assessment of ionosphere/thermosphere models: NmF2, hmF2, and vertical drift using ground-based observations. *Space Weather*, *9*, S12003. <https://doi.org/10.1029/2011SW000727>
- Smith, C. W., Acuña, M. H., Burlaga, L. F., L'Heureux, J., Ness, N. F., & Scheifele, J. (1998). The ACE magnetic fields experiment. *Space Science Review*, *86*, 613. <https://doi.org/10.1023/A:1005092216668>

- Spence, H. E., Reeves, G. D., Baker, D. N., Blake, J. B., Bolton, M., Bourdarie, S., et al. (2013). Science goals and overview of the Energetic Particle, Composition, and Thermal Plasma (ECT) suite on NASA's Radiation Belt Storm Probes (RBSP) mission. *Space Science Reviews*, *179*, 311–336. <https://doi.org/10.1007/s11214-013-0007-5>
- Taktakishvili, A., Kuznetsova, M. M., MacNeice, P., Hesse, M., Rastätter, L., Pulkkinen, A., et al. (2009). Validation of the coronal mass ejection predictions at the Earth orbit estimated by ENLIL heliosphere cone model. *Space Weather*, *7*, S03004. <https://doi.org/10.1029/2008SW000448>
- Verbeke, C., Mays, M. L., Temmer, M., Bingham, S., Steenburgh, R., Dumbović, M., et al. (2019). Benchmarking CME arrival times and impact: Progress on metadata, metrics, and events. *Space Weather*, *17*, 6–26. <https://doi.org/10.1029/2018SW002046>
- Wold, A. M., Mays, L. M., Taktakishvili, A., Jian, L. K., Odstrcil, D., & MacNeice, P. (2018). Verification of real-time WSA-ENLIL+Cone simulations of CME arrival-time at the CCMC from 2010 to 2016. *Journal of Space Weather and Space Climate*, *8*, A17. <https://doi.org/10.1051/swsc/2018005>
- Wolff, J. K., Harrold, M., Fowler, T., Gotway, J. H., Nance, L., & Brown, B. G. (2014). Beyond the basics: Evaluating model-based precipitation forecasts using traditional, spatial, and object-based methods. *Weather Forecasting*, *29*, 1451–1472. <https://doi.org/10.1175/WAF-D-13-00135.1>
- Yu, Y., Jordanova, V., Zaharia, S., Koller, J., Zhang, J., & Kistler, L. M. (2011). Validation study of the magnetically self-consistent inner magnetosphere model RAM-SCB. *Journal of Geophysical Research*, *117*, A03222. <https://doi.org/10.1029/2011JA017321>
- Yu, Y., Rastätter, L., Jordanova, V. K., Zheng, Y., Engel, M., Fok, M.-C., & Kuznetsova, M. M. (2019). Initial results from the GEM challenge on the spacecraft surface charging environment. *Space Weather*, *17*, 299–312. <https://doi.org/10.1029/2018SW002031>