Geospace Environment Modeling 2008–2009 Challenge: Geosynchronous magnetic field

L. Rastätter,[1] M. M. Kuznetsova,[1] A. Vapirev,[2] A. Ridley,[3] M. Wiltberger,[4] A. Pulkkinen,[1,5] M. Hesse,[1] and H. J. Singer[6]

Received 7 September 2010; revised 14 January 2011; accepted 15 March 2011; published 28 April 2011.

[1] In this paper the metrics-based results of the inner magnetospheric magnetic field part of the 2008–2009 GEM Metrics Challenge are reported. The Metrics Challenge asked modelers to submit results for four geomagnetic storm events and five different types of observations that can be modeled by statistical or climatological or physics-based (e.g., MHD) models of the magnetosphere-ionosphere system. We present the results of 12 model settings that were run at the Community Coordinated Modeling Center and at the institutions of various modelers for these events. To measure the performance of each of the models against the observations, we use direct comparisons between the strength of the measured magnetic field (B), the sine of the elevation angle Θ_{xz} (τ), and the spectral power of fluctuations for both quantities. We find that model rankings vary widely by type of variable and skill score used. None of the models consistently performs best for all events. We find that empirical models perform well for weak storm events, and physics-based (magnetohydrodynamic) models are better for strong storm events. Within a series of runs of the same model we find that higher resolution may not always improve results unless more physics of the inner magnetosphere, such as the kinetic description of the ring current, is included.

Citation: Rastätter, L., M. M. Kuznetsova, A. Vapirev, A. Ridley, M. Wiltberger, A. Pulkkinen, M. Hesse, and H. J. Singer (2011), Geospace Environment Modeling 2008–2009 Challenge: Geosynchronous magnetic field, *Space Weather*, *9*, S04005, doi:10.1029/2010SW000617.

1. Introduction

[2] In recent years, the quantitative assessment of the performance of various statistical and physics-based models has become more important to serve the space weather community, with an increasing number of applications that specify and predict space weather conditions. With quantifiable metrics, users of space weather modeling products will be able to better understand the strengths and weaknesses of each modeling approach and select the approach best suited for their application. In addition to serving the user, modelers gain insight into how different modeling parameters influence the performance of a given model and how different versions of a model are improving over time.

[3] A metrics challenge for state-of-the-art global magnetospheric space weather models has been discussed for years in the Geospace Environment Modeling (GEM) community. The GEM Global Geospace Circulation Modeling (GGCM) Metrics and Validation Focus Group organized a Modeling Challenge to focus on the dynamics of the inner magnetosphere and ground magnetic field perturbations. The 2008–2009 challenge was defined at the 2008 GEM workshop in Midway, Utah, and was broadly announced in September 2008. Model result submissions were accepted through the Community Coordinated Modeling Center (CCMC) and submissions received through 1 September 2009 are included in this paper. Besides the online submission system, an online model comparison tool is available on the CCMC web site to compare existing submissions to observations.

[4] This new activity follows a series of earlier GEM challenges [*Lyons*, 1998; *Birn et al.*, 2001; *Raeder and Maynard*, 2001; *Ridley et al.*, 2002] but extends the focus on ionospheric convection events and isolated substorms to geomagnetic storms and observations on the ground

[1]Community Coordinated Modeling Center, Space Weather Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.
[2]Space Science Center, University of New Hampshire, Durham, New Hampshire, USA.
[3]Department of Atmospheric, Oceanic, and Space Sciences, University of Michigan, Ann Arbor, Michigan, USA.
[4]High Altitude Observatory, National Center for Atmospheric Research, Boulder, Colorado, USA.
[5]IACS, Catholic University of America, Washington, D. C., USA.
[6]Space Weather Prediction Center, National Centers for Environmental Prediction, NOAA National Weather Service, Boulder, Colorado, USA.

Table 1. Event Numbers With Dates, Minimum Dst, Maximum Kp, and GOES Satellites

Event	Date and UT Time	Min Dst (nT)	Max Kp	GOES Satellites[a]
1	29 Oct. 2003, 0600 to 30 Oct. 2003, 0600	−353	9	10, 12
2	14 Dec. 2006, 1130 to 16 Dec. 2006, 0000	−139	8	11, 12
3	31 Aug. 2001, 0000 to 1 Sep. 2001, 0000	−40	4	10, 8
4	31 Aug. 2005, 0930 to 1 Sep. 2005, 1200	−131	7	10, 12

[a]Format of values is asterisk, diamond. Symbols used in Figures 5–12 with prediction efficiencies and log-spectral distance scores are listed.

and in geosynchronous orbit. This study is based on four full storm events that contain a large range of geomagnetic states. The goals of this challenge are to evaluate differences between the available modeling approaches, the effect of the inclusion of kinetic modeling of the ring current and model resolution. This challenge is the first in a series of challenges that can be used by anyone to track the performance measures as models improve. The ongoing comparison of observations and models will also encourage collaborations between modelers and data analysts.

[5] After a preliminary analysis of a subset of the modeling submissions for both applications (magnetic perturbations on the ground and in geosynchronous orbit) [*Pulkkinen et al.*, 2010], we are now presenting the full set of model submissions for magnetic fields at geosynchronous orbit in this paper. The analysis of ground perturbations with a larger set of models has been described separately [*Pulkkinen et al.*, 2011].

[6] The scope of this paper is to report on the overall performance of the submitted model setups by using model outputs obtained at applicable satellites in the inner magnetosphere without actually analyzing the three-dimensional output data (if available) of each individual model simulation. More detailed scientific analysis (e.g., of the effect of various model parameter settings on the ability to reproduce observed spatial and temporal features) is best carried out individually for each model in future papers. The metrics evaluation tool available online at the CCMC will facilitate such an analysis as more model runs are submitted in future challenges.

2. Setup of the Challenge

[7] Four geospace events were selected. Two represent highly disturbed times, event 1 from 29 October 2003, 0600 UT to 30 October 2003, 0600 UT, known as part of the "Halloween Storm" and event 2 from 14 December 2006, 1130 UT to 16 December 2006, 0000 UT, known as the "AGU Storm." The other two events represent quieter times, which are event 3 from 31 August 2001, 0000 UT to 1 September 2001, 0000 UT and event 4 from 31 August 2005, 0930 UT to 1 September 2005, 1200 UT. All events, with their start and end dates and times, minimum Dst, and maximum Kp index values, are listed in Table 1.

[8] For each of the events the solar wind magnetic field and plasma parameters obtained by the MAG and SWEPAM instruments on the Advanced Composition Explorer (ACE) satellite are shown in Figure 1. All events were covered by the ACE measurements, except event 1 in Figure 1a, the Halloween Storm, for which plasma velocity data could be reconstructed only with low time resolution [*Skoug et al.*, 2004]. Plasma density data were constructed from the Plasma Wave Instrument on the Geotail spacecraft. Events 1 and 2 (shown in Figures 1a and 1b, respectively) are large CME-related storm events whereas events 3 and 4 (shown in Figures 1c and 1d, respectively) are less active periods. Each of Figures 1a–1d has eight panels showing the plasma density in atomic mass units per cm^3, temperature in K, GSM velocity components in km/s and GSM magnetic field components in nT. ACE is located near the first Lagrange point, 230 (±5) Earth radii away from Earth, near the Sun-Earth line. During event 1 on 29 October 2003, Geotail was located between 15 and 30 Earth radii in X_{GSE}, upstream of the bow shock on the duskside (between 5 and 15 R_E in X_{GSE}). Solar wind data have been propagated with the solar wind velocity to the sunward boundary of the magnetospheric modeling region (about $33R_E$) or to Earth (for statistical Tsyganenko models). The magnetic field component B_x is set to zero for all magnetohydrodynamic simulations to preserve $\nabla \cdot \mathbf{B} = 0$.

[9] The GEM challenge strives to evaluate, for each event, the models' performance by comparing the specifications with observations of the following geospace parameters: (1) magnetic field at geosynchronous orbit, (2), magnetopause crossings by geosynchronous satellites, (3) plasma pressure or density/temperature at geosynchronous orbit, (4) ground magnetic field perturbations, and (5) Dst index.

[10] The focus of this study is to evaluate how well models represent the magnetic field at geosynchronous orbit as observed by the NOAA Geosynchronous Operational Environmental Satellites (GOES). For each event studied there are two operational satellites (event 1, GOES 10 and GOES 12; event 2, GOES 11 and GOES 12; event 3, GOES 8 and GOES 10; and event 4, GOES 10 and GOES 12; see Table 1). For all events (except event 4), GOES magnetic field data were available from CDAWeb. For event 4, data were reprocessed and directly obtained from NOAA. For all events magnetic field data are available with only small gaps, which were filled in by interpolation. GOES satellites are located at geosynchronous

Figure 1. Solar wind bulk plasma and magnetic field observations from ACE and Geotail for the four storm events listed in Table 1: (a) event 1, (b) event 2, (c) event 3, and (d) event 4.

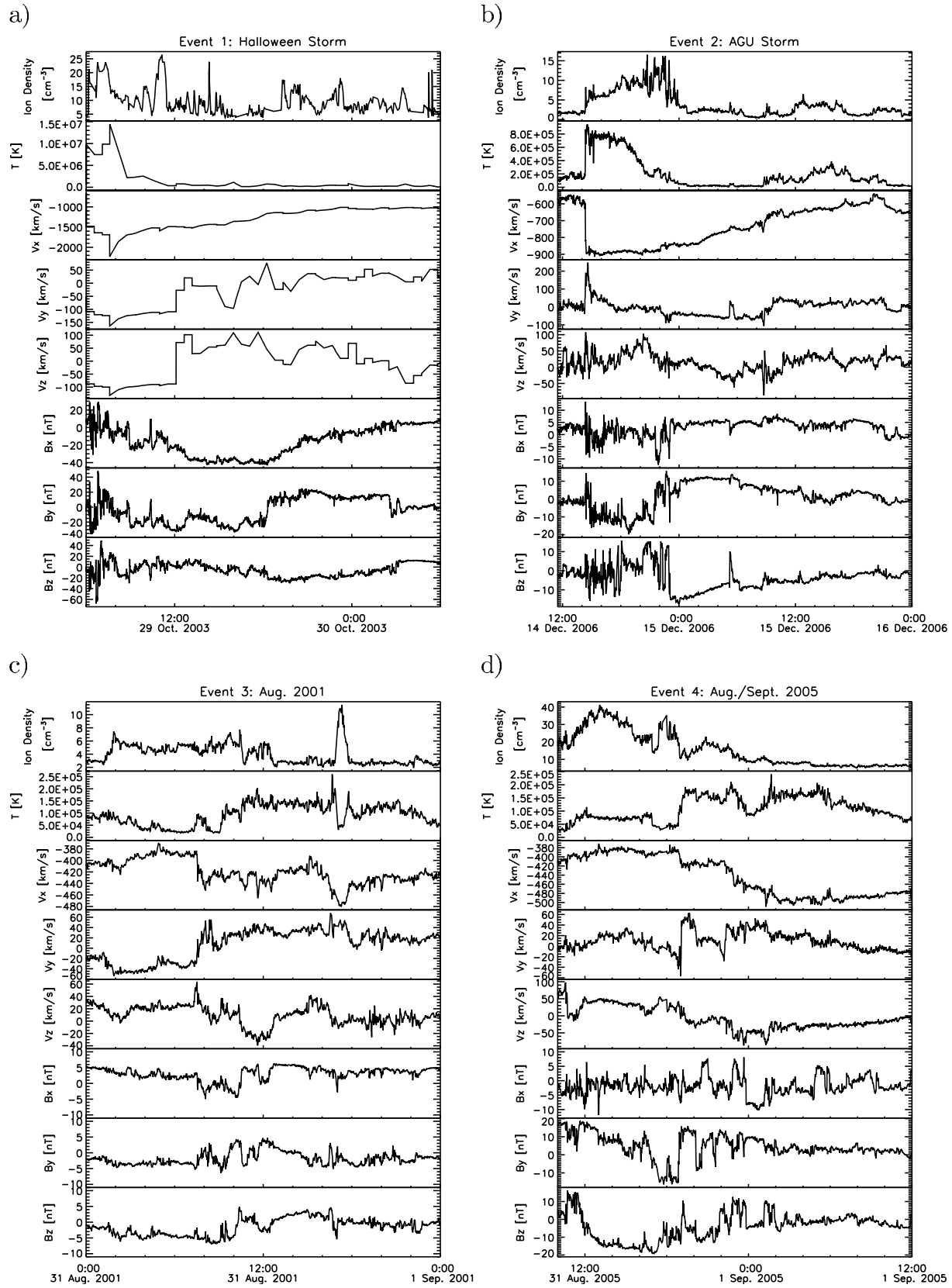
a)
Event 1: Halloween Storm
Ion Density [cm⁻³]
T [K]
Vx [km/s]
Vy [km/s]
Vz [km/s]
Bx [nT]
By [nT]
Bz [nT]
12:00 29 Oct. 2003
0:00 30 Oct. 2003
b)
Event 2: AGU Storm
12:00 14 Dec. 2006
0:00 15 Dec. 2006
12:00 15 Dec. 2006
0:00 16 Dec. 2006
c)
Event 3: Aug. 2001
0:00 31 Aug. 2001
12:00 31 Aug. 2001
0:00 1 Sep. 2001
d)
Event 4: Aug./Sept. 2005
12:00 31 Aug. 2005
0:00 1 Sep. 2005
12:00 1 Sep. 2005

Figure 1

orbit (distance form the Earth's center: 6.62 R_E). Two satellites are operational at any time; one satellite is located over the eastern United States (about 75 degrees west longitude), the other in the west (about 135 degrees west longitude).

3. Models Used

[11] Various modelers have tested their own model with GOES observations. Submissions to this Challenge include runs with the Lyon-Fedder-Mobarry model component of the Coupled Magnetosphere, Ionosphere, and Thermosphere model (CMIT-LFM) [*Wiltberger et al.*, 2000; *Lyon et al.*, 2004; *Wiltberger et al.*, 2004], OpenGGCM [*Raeder et al.*, 2001] and SWMF [*Tóth et al.*, 2005]. Models run at the CCMC include SWMF/BATSRUS, OpenGGCM and the statistical magnetic field models Tsyganenko-96 [*Tsyganenko*, 1996] and Tsyganenko-2004 (also known as Tsyganenko-Sitnov 2004) [*Tsyganenko and Sitnov*, 2005]. With the modeling and metrics challenge one can for the first time compare the performance of major magnetosphere MHD models and empirical models under the same conditions: all models are run for the same events and the same analysis tools can be applied to compute performance skill scores from each run. In addition to the model runs performed at the CCMC, modelers were invited to submit their own results to CCMC for metrics-based analysis. It is important to note that this metrics study is the first in a series of studies that will use the same event data. Future versions of each model may be run with the same inputs and state-of-the-art model settings (reflecting advances in physical understanding, numerical methods and computational resources) will be applied. Over time, model performance can be followed to track midterm to long-term trends.

3.1. Space Weather Modeling Framework

[12] The most commonly requested model of the magnetosphere at the CCMC is the Space Weather Modeling Framework (SWMF), consisting of the magnetosphere MHD component GM/BATSRUS [*Powell et al.*, 1999] which is coupled to an ionospheric potential solver IE/Ridley_Serial [*Ridley et al.*, 2004]. Optional is the inclusion of Version 2 (within SWMF) of the Rice Convection Model as an inner magnetospheric component IM/RCM2 [*Sazykin et al.*, 2002]. The SWMF framework is described in detail by *Tóth et al.* [2005]. The BATSRUS component can track satellites in GSM coordinates throughout the magnetosphere.

[13] We performed five runs for each event with SWMF, including the setup for real-time simulation (755,000 grid cells with $1/4R_E$ resolution), 2 million grid cells with $1/4R_E$ resolution (with and without RCM), and 3 million grid cells with $1/8R_E$ resolution (with and without RCM). The model uses an adaptive grid composed of rectangular blocks with varying resolution. The finest blocks with $0.25R_E$ or $0.125R_E$ resolution are concentrated around the Earth, in the dayside (up to the magnetopause), and in parts of the magnetotail along the x axis. The Earth's dipole orientation in GSM coordinates is updated in time.

3.2. OpenGGCM

[14] The Open Geospace General Circulation Model (OpenGGCM) is a magnetospheric MHD model that is coupled to the ionospheric CTIM model as described by *Raeder et al.* [2001]. The magnetospheric grid is a single block with cells of varying size and shape (each coordinate has varying grid spacing) in GSE coordinates. Low-resolution runs used 3 million cells with finest resolution near Earth of $0.3R_E$ and higher-resolution runs 6.5 million cells with a minimum resolution near Earth of $0.25R_E$.

[15] OpenGGCM simulations are performed in GSE coordinates and have a Earth dipole that is fixed in time. For long event runs (a day or longer), the dipole tilt is set so it matches the average tilt in the x-z plane and y-z plane during the event. The model's stretched cartesian grid resolves the dayside and near-Earth region fairly well and also resolves the tail region well in the y and z directions. To better represent the magnetic field at geosynchronous orbit, we corrected the magnetic field written by the model at the satellite locations by subtracting the Earth's dipole field with the fixed axis that was used in the model and adding the dipole field with time-varying axis orientation.

3.3. Coupled Magnetosphere, Ionosphere, and Thermosphere

[16] CMIT-LFM simulations [*Lyon et al.*, 2004; *Wiltberger et al.*, 2004] are performed in SM coordinates. The LFM magnetosphere grid is a distorted spherical grid. The magnetopause and bow shock region in the dayside are resolved well in the radial direction. Nightside features, such as the magnetotail, are resolved less.

3.4. Model Summary

[17] Magnetospheric MHD models differ in design but essentially solve the same set of ideal MHD equations. The magnetosphere module in each model is coupled to an ionospheric potential solver that uses a conductance which is driven by solar irradiation (represented by the solar 10.7 cm flux parameter) and magnetospheric field-aligned currents that have been mapped into the ionosphere. High spatial-order numerical schemes used are matched with shock-capturing limiters and, in the case of OpenGGCM, an additional artificial resistivity which ensures numerical stability. In contrast to the up to eighth-order spatial discretization in LFM and up to fourth-order discretization in OpenGGCM, SMWF simulations that were run at CCMC employ a second-order Rusanov scheme that is fairly diffusive and that creates a smoother overall solution. One SWMF run was performed with a less diffusive second-order Sokolov scheme [*Sokolov et al.*, 1999]. Recently, *Ridley et al.* [2010] examined the effect of numerical schemes on the SWMF, illustrating that while less diffusive schemes allow sharper gradients to build in the magnetosphere, they do not necessarily provide the best data-model comparisons.

Table 2. Model Run Settings Used in the Challenge[a]

Model Description	Identifier
LFM, 160,000 cells, min. res. 0.15 R_e radial	1_LFM
OpenGGCM v3.1 coupled to CTIM, 3 million cells, min. res. 0.3 R_E	1_OPGM
OpenGGCM v3.1 coupled to CTIM, 6.5 million cells, min. res. 0.25 R_E (CCMC)	2_OPGM
OpenGGCM v3.1 coupled to CTIM, 6.5 million cells, min. res. 0.25 R_E (UNH)	3_OPGM
SWMF v7.73, BATS-R-US, 2 million cells, min. res. 0.25 R_E	1_SWMF
SWMF v7.73, BATS-R-US, 700,000 cells, min. res. 0.25 R_E	2_SWMF
SWMF v8.01, BATS-R-US coupled to RCM, 2 million cells, min. res. 0.25 R_E	3_SWMF
SWMF v8.01, BATS-R-US, 3 million cells, min. res. 0.125 R_E	4_SWMF
SWMF v8.01, BATS-R-US coupled to RCM, 3 million cells, min. res. 0.125 R_E	5_SWMF
SWMF v20090403, BATS-R-US coupled to RCM, 900,000 cells, min. res. 0.25 R_E	6_SWMF
Tsyganenko-96 model	1_T96
Tsyganenko-Sitnov 2004 model	1_T04

[a]Simulation runs of 1_LFM, 1_OPGM, 2_OPGM, 1_SWMF, 3_SWMF, 4_SWMF, and 5_SWMF can be requested through the "Request A Model Run" capability at CCMC.

[18] The SWMF and OpenGGCM MHD models have the capability to track satellites during the execution of the simulations. For each of the two operational GOES satellites per event, satellite trajectory data in the model coordinates (GSM for BATSRUS, GSE for OpenGGCM) were obtained from the NASA Satellite Situation Center ("Locator Tabular" web interface) and fed into the models. The models then output the MHD variables at the desired time resolution for each satellite for comparison with the observations.

[19] In addition to MHD models, we ran two versions of the Tsyganenko magnetic field model: T96 [*Tsyganenko*, 1995, 1996] and T04 [*Tsyganenko and Sitnov*, 2005]. The Coupled Magnetosphere, Ionosphere and Thermosphere (CMIT) model with LFM magnetosphere and MIX ionosphere [*Wiltberger et al.*, 2004; *Merkin et al.*, 2007] was exclusively run by the modelers. Several models were run with different grid and spatial resolution settings (see Table 2).

4. Analysis and Skill Scores

[20] The magnetic field time series can be compared in several ways, each having strengths and weaknesses that will be pointed out in the following sections and section 6.

4.1. Variables Compared

[21] For the metrics evaluation, one needs to select a time series of a variable that can be obtained both from the observations and model results.

[22] With magnetic field measurement, one can look at the individual components of the magnetic field, the magnitude or orientation angles of the vector. In this work we concentrate on two parameters derived from the magnetic field measurements and model specifications: magnetic field strength (B) and magnetic elevation (τ).

4.1.1. Magnetic Field Strength B

[23] The total magnetic field strength

$$B_{\mathrm{dip}} \sim -31100\mathrm{nT}\left(\frac{R_E}{r}\right)^3\sqrt{1+3\sin^2(\lambda)} \quad (1)$$

averages about 105 nT at geomagnetic latitude $\lambda = 0$ and geosynchronous orbit $r = 6.62R_E$ and rarely drops below 50 nT. Changes of B beyond smooth diurnal variations from the dipole values indicate such features as dipolarizations in the nightside and compressions in the dayside of the magnetosphere. To ensure that models and observations are only compared when the GOES satellites are inside the magnetosphere, only time periods when the north-south component of the magnetic field in GSM coordinates (B_z) remains above zero are considered. Since B_z almost never dips below zero unless the satellite crosses the magnetopause in the dayside [see, e.g., *Rufenach et al.*, 1989; *Welling and Ridley*, 2010], this will exclude time periods when the observing satellite is in fact (or is modeled to be) outside the magnetopause. Possibly, one could look at B_z as it is the major magnetic field component at geosynchronous orbit, but the total magnetic field strength is better suited to evaluate magnetospheric pressure balance and magnetopause standoff distance.

4.1.2. Magnetic Elevation τ

[24] The magnetic elevation angle $\Theta_{xz} = \arctan(B_z, B_x)$ in GSM coordinates provides a measure of magnetic field inclination or field line stretching parallel to the noon-midnight meridional plane. The largest variations in the inclination angles are usually observed in the nightside near local midnight where magnetotail stretching reduces the elevation angle and dipolarization increases the angle. The angle spans the range from –180 degrees (for $B_x < 0$, $B_z \nearrow 0$ and $|B_x|/|B_z| \gg 1$) to +180 degrees (for $B_x < 0$, $B_z \searrow 0$, $|B_x|/|B_z| \gg 1$). To avoid the discontinuity when B_z passes from negative to positive or vice versa in the presence of negative B_x values, we use

$$\tau = \sin(\Theta_{xz}) = \frac{B_z}{\sqrt{B_x^2 + B_z^2}} \quad (2)$$

as the primary variable that measures elevation. The variable ranges from –1 (southward) to +1 (northward magnetic field). Here, negative elevations indicate dayside magnetopause crossings. Increasing elevations ($\tau \rightarrow +1$) indicate dipolarization of the inner magnetosphere on the nightside. For the calculation of variances and skill scores, only time periods with positive values of τ are considered. The use of the elevation also eliminates effects of offsets of the magnetic field strength that may occur in modeling results.

[25] Figure 2 shows for each event (Figures 2a–2d) the magnetic field observations (B, B_x, B_y, B_z), elevation angle Θ_{xz} and τ for two satellites in a total of six panels. Magnetic field strength B_z is usually positive, except when the

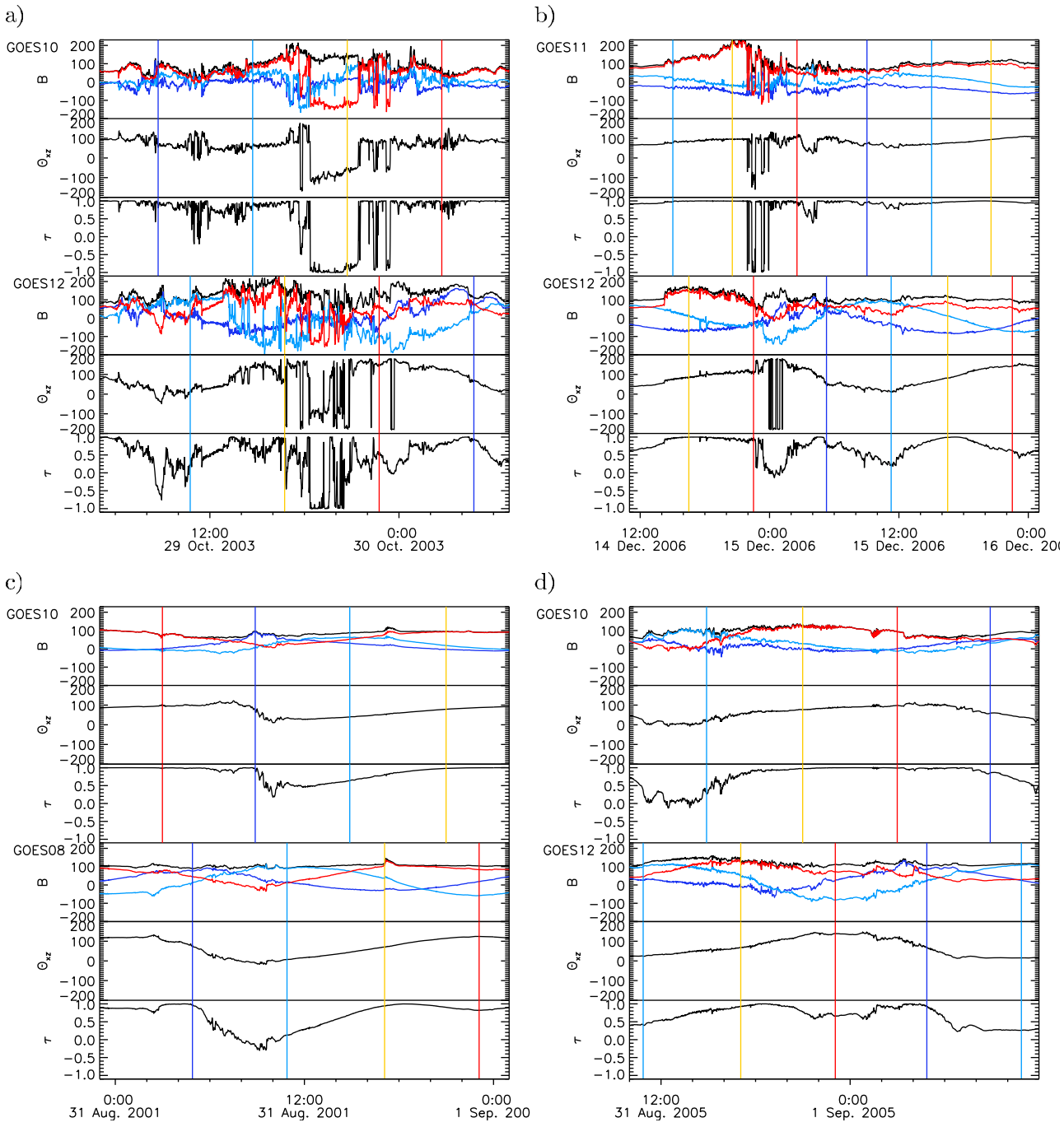

Figure 2. (a–d) The three top panels belong to one GOES satellite, and the other three belong to the other GOES satellite in operation at each event. Observed GOES magnetic field (black, B; purple, B_x; blue, B_y; and red, B_z), elevation angle Θ_{xz}, and τ are shown for each satellite. Vertical lines indicate midnight (purple), dawn (blue), noon (yellow), and dusk (red) local time for each spacecraft.

satellite crosses the magnetopause in the dayside (i.e., the time between the blue and red vertical lines indicating dawn and dusk local time, respectively) during a strong storm or when strong magnetic disturbances occur in the nightside, such as seen at GOES 12 in event 1 between 0800 UT and 0900 UT on 29 October 2003.

4.2. Types of Skill Scores

[26] As a companion paper to *Pulkkinen et al.* [2011] we use two skill scores that are applicable to this study, prediction efficiencies (*PE*) and log-spectral distance (M_s).

4.2.1. Prediction Efficiency *PE*

[27] The variables explained in sections 4.1.1 and 4.1.2 can be evaluated by computing a prediction efficiency (PE), defined for a discrete time series as

$$PE = 1 - \frac{\langle (x_{mod} - x_{obs})^2 \rangle}{\sigma_{obs}^2} \quad (3)$$

with $\langle \; .. \; \rangle$ denoting the arithmetic mean, x_{obs} the observation, x_{mod} the modeled signal, and σ_{obs}^2 the variance of the observed signal. $PE = 1$ indicates perfect model performance and $PE = 0$ indicates performance comparable to predicting the arithmetic mean of the observed signal. PE can reach unlimited negative values.

4.2.2. Log-Spectral Distance M_s

[28] The spectral power determines the level of disturbance on different timescales that can be produced by the models in comparison to the observed level of fluctuations. Solar wind disturbances are a source of spectral power in the magnetosphere-ionosphere system, and the latter may act as a low-pass filter that exhibits a different spectral distribution in the magnetosphere and the ionosphere than in the solar wind driver.

[29] The analysis of the log-spectral distance evaluates the spectral distribution of fluctuations in a given spectral range by computing a single number that measures the discrepancy between the observed spectral distribution from that obtained by a model. The comparison of spectral distributions in the model outputs compared to the observations (and solar wind inputs) indicates how well a model preserves activity levels in various frequency ranges. A model would perform perfectly if the spectral distribution in the observations matched the modeled spectrum.

[30] To compute the log-spectral distance, the logarithm of the ratio of the spectral power of the observed ($|\tilde{x}_{obs}|$) and modeled variable ($|\tilde{x}_{mod}|$):

$$m_s = \log\left[\frac{|\tilde{x}_{mod}|}{|\tilde{x}_{obs}|}\right] \quad (4)$$

is calculated and then summed over the N frequencies f to yield the log-spectral distance M_s:

$$M_s = \sqrt{\frac{1}{N}\sum_f m_s^2} \quad (5)$$

[31] The score M_s is equal or larger than zero. $M_s = 0$ is a perfect score. To perform the spectral analysis, 2 h length windows are selected from the 1 min data and model results, yielding a Fourier spectrum for periods between 2 min ($f_1 = 1/120$ Hz) and 60 min ($f_2 = 1/3600$ Hz). Windows with negative B_z values in either observed or modeled data are not included to avoid high-frequency signals introduced by magnetopause crossings. A 75% overlap between adjacent windows is allowed to capture data and model outputs in 30 min increments while avoiding the loss of an entire 2 h time period as the result of a single rejected window. Spectra from all valid windows are averaged to form the spectra from the observation data ($\tilde{x}_{obs}$) and model outputs ($\tilde{x}_{mod}$).

5. Results

5.1. Magnetic Field *B*

[32] Figures 3 and 4 show for each event the observations of B (black line) together with the traces for each model in the top plot and of τ in the bottom plot for each satellite. Figure 3a features event 1, and Figure 3b features event 2. Figure 4a features event 3, and Figure 4b features event 4. The top two panels feature one satellite and the bottom two panels the other satellite in operation for each event.

[33] In addition to obtaining the overall prediction efficiencies and log-spectral distances for all events and local times, we separated the comparisons into strong events (events 1 and 2) and weak events (events 3 and 4) and as well as dayside (satellite positions with $X_{GSE} > 0$) and nightside ($X_{GSE} < 0$) during the events.

[34] Figure 5 shows the ranking of the models based on the average for all events and all local times for variable B for the prediction efficiency in Figure 5a and log-spectral distance in Figure 5b. The best ranking model is shown at the left in each panel. The prediction efficiencies vary widely with most models performing rather poorly for the weak events (green and red symbols) and better for strong events (black and blue symbols near the top of the graph). Log-spectral distances vary by about a factor of 3 (about half a decade in the logarithmic vertical axis of the plot) for most models between all the events. Compared to the prediction efficiencies in Figure 5a, there is no apparent distinction between stronger and weaker events. Each event is represented by two symbols (asterisk, diamond), representing the two GOES satellites in operation for each event. The two scores obtained for each event are similar but are influenced by the difference in local time between the two satellites.

[35] To further examine the distinction between prediction efficiencies for weak and strong events, we show strong and weak storm events separately and compute the average skill scores for each subset of events. Figure 6 shows the rankings for weak events. We note that the two weak events show a clear separation in terms of

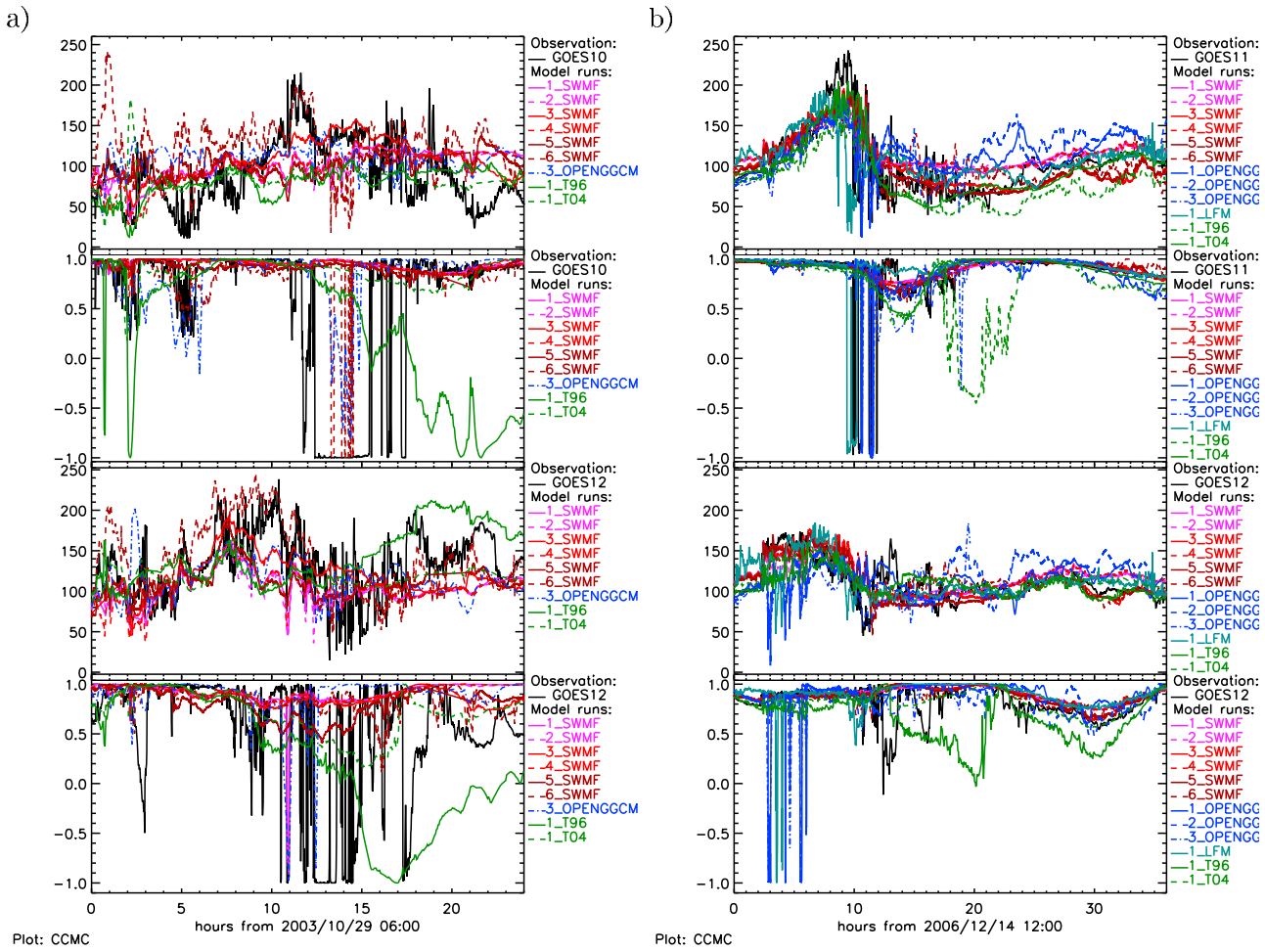

Figure 3. (a) Event 1 and (b) event 2: the top two panels belong to one GOES satellite, and the bottom two belong to the other GOES satellite in operation at each event. Magnetic field strength B (observation in black, model results in color) are shown in the first and third panels from the top, and τ is shown in the second and fourth panels. Two panels (one for B and one for τ) belong to each of the two GOES satellites in operation during each event as listed in Table 1.

prediction efficiency between the Tsyganenko, LFM and SWMF runs as opposed to the OpenGGCM runs, with the latter showing large negative values for the prediction efficiencies (weak rankings) and fairly low values of the log-spectral distances (good rankings). Note that run 6_SWMF is in the middle in the PE ranking and first in the M_s ranking but was run only for event 3. Prediction efficiencies span a wide range (–34 to +1). All models perform worse for event 3 than for event 4 in terms of prediction efficiency but the result is mixed with respect to the log-spectral distance.

[36] In Figure 7, the models are ranked for strong storm events. We notice that in general, most OpenGGCM runs trail in PE and perform better for M_s. The range of PE values (y axis) has shrunk to a narrow band ($PE > -1.1$) but the rankings (order of model runs along the x axis) in general are very similar to the ranking for the weak storms and the overall ranking. Note that Figure 7 only includes 1_LMF and 1_OPGM for event 2 and did not include event 1 where all models perform worse. Both runs would rank much weaker (rank 9 and 11, respectively) if we considered only event 2 (blue symbols). With regard to log-spectral distance in Figure 7b, 1_LFM and OpenGGCM runs show lower log-spectral distances than SWMF and Tsyganenko, consistent with the observation for weak events. The fact that all models perform worse for event 1 compared to event 2 (for both, PE and M_s) may be caused by the reconstructed solar wind plasma data (velocity and density, in Figure 1a which may lack some of the high-frequency variations seen in the magnetic field data. The prediction efficiencies for strong events are vastly better than those for weak events and are confined to a narrow band between –1 and 0.5.

[37] Tsyganenko-Sitnov-2004 shows a large improvement over Tsyganenko-96 in terms of prediction efficiency whereas the performance in terms of log-spectral distance

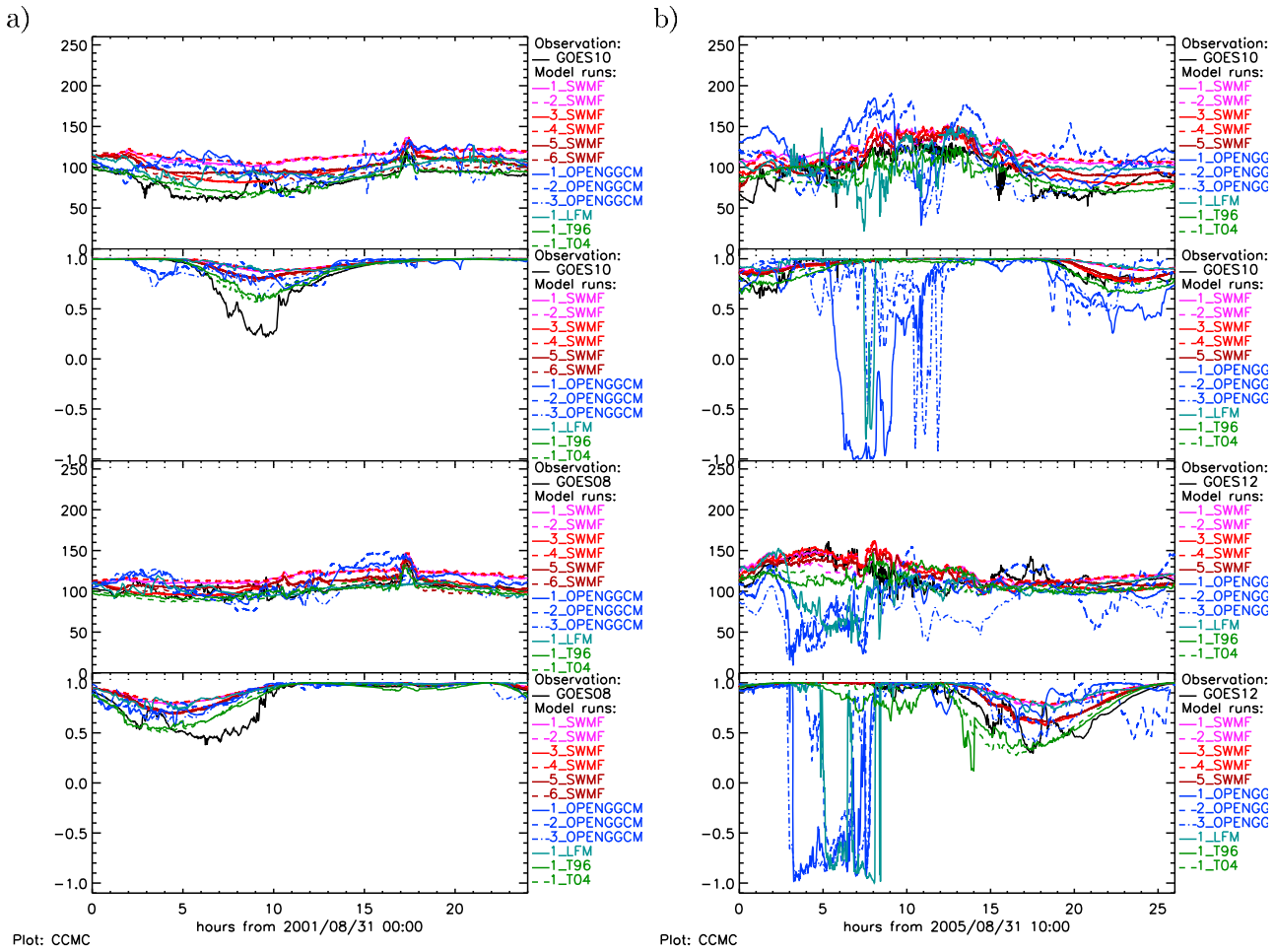

Figure 4. (a) Event 3 and (b) event 4 in the same format as in Figure 3.

is worse. Clearly, the model is optimized to perform well when representing magnetic field values, on average, as opposed to reproducing short-term fluctuations of the field values.

[38] In the series of SWMF runs, we note that 3_SWMF performs best overall (best in PE, second in terms of M_S) followed by 6_SWMF (third in PE, first in M_S) and 5_SWMF (second in PE, third in M_S). The consistently good performance of the three SWMF runs with RCM inner magnetosphere indicates that these model setups improve how the kinetic physics of the inner magnetosphere is represented in an MHD framework. Run 3_SWMF performs slightly better compared to 5_SWMF, which had finer resolution and more cells, and much better than 4_SWMF (finer resolution without RCM). Run 6_SWMF performs fairly well in spite of its much lower number of grid cells at $1/4R_E$ resolution. 6_SWMF is run using the less diffusive Sokolov discretization scheme and with β = 1.2 in the MC3 limiter. The other SWMF runs (including 5_SWMF) used the Rusanov scheme and the MC3 limiter with β = 1.0 (equivalent to the "minmod" limiter as described by *Tóth et al.* [2006]). Overall, higher spatial resolution does not necessarily improve model performance of the SWMF (BATSRUS magnetosphere) model, as was also noted by *Welling and Ridley* [2010]. This run was submitted by the SWMF developers and used a less diffusive discretization scheme (Sokolov) and a sharper limiter (MC3 with β = 1.3) than the other SWMF runs (Rusanov scheme, "minmod" limiter β = 1.0). These settings that the modelers used come at the price of reduced model stability and thus are not commonly used at the CCMC.

[39] The series of OpenGGCM runs shows no clear distinction for different settings. All three run settings are always close together in the rankings. The OpenGGCM runs do better in the log-spectral distance rankings than for prediction efficiency.

[40] Figures 8 and 9 show rankings for the prediction efficiencies and log-spectral distances, respectively. In deriving the rankings, only time windows that are fully in the nightside (satellite positions $X_{GSM} < 0$; Figures 8a and 9a), and the dayside ($X_{GSM} > 0$; Figures 8b and 9b)

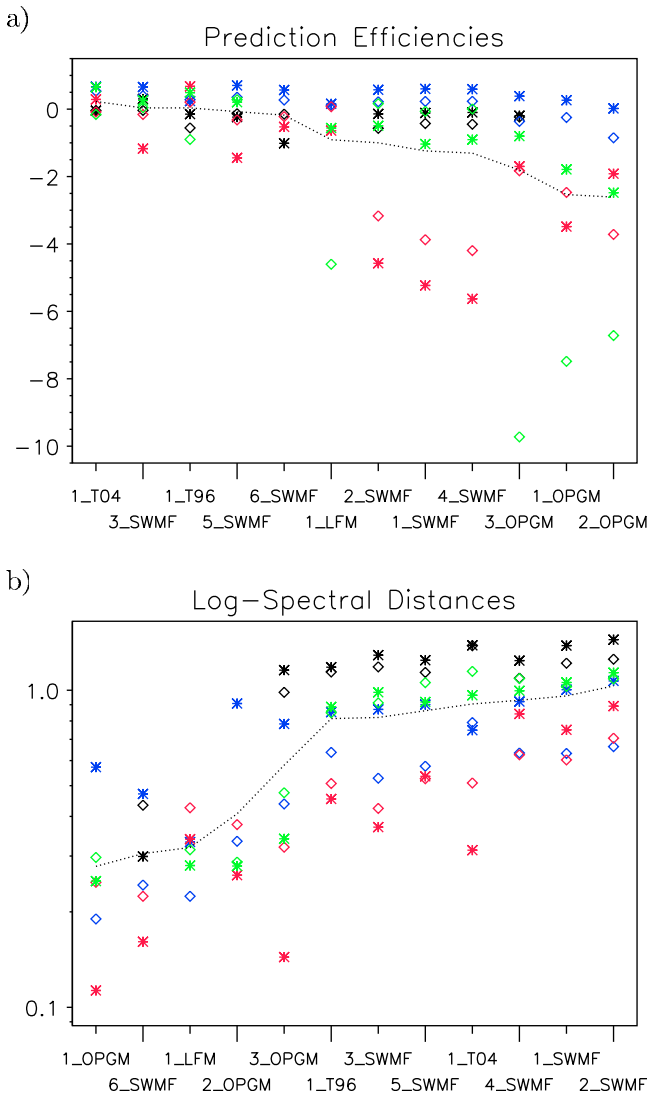

Figure 5. Ranking of model runs for B according to (a) prediction efficiency and (b) log-spectral distance averaged over events and GOES satellites. The differently colored symbols refer to the events: black, event 1; blue, event 2; red, event 3; and green, event 4. The asterisk and diamond symbols show the scores obtained for the locations of two GOES satellites in operation at each event (see Table 1). The dashed line is the average score that was used to sort the models in the plots.

have been considered. Although the skill scores may differ considerably from the scores for the whole events (at least a full 24 h), we find only slight modifications of the rankings for model runs between nightside and dayside compared to the whole events. In terms of prediction efficiencies, we find that most scores are lower when considering subsets of each event for the nightside or the dayside. The reason for this is a reduced denominator (σ^2_{obs} in equation (3)) that results from comparing data with an average taken from a shorter time period. Averages from the nightside or dayside are different due to regular diurnal variations of the magnetic field strength. Modeled data often do not fully follow the diurnal variation seen in the observed satellite data and thus show larger deviations ($\langle\langle(x_{\text{mod}} - x_{\text{obs}})^2\rangle\rangle$), resulting in lower scores.

5.2. Magnetic Elevation τ

[41] Computing the elevation angle can reduce modeling bias in terms of magnetic field strength which can come about by missing inner magnetospheric physics in first-principles models. If biases are present in modeling results, we may see higher values in prediction efficiencies

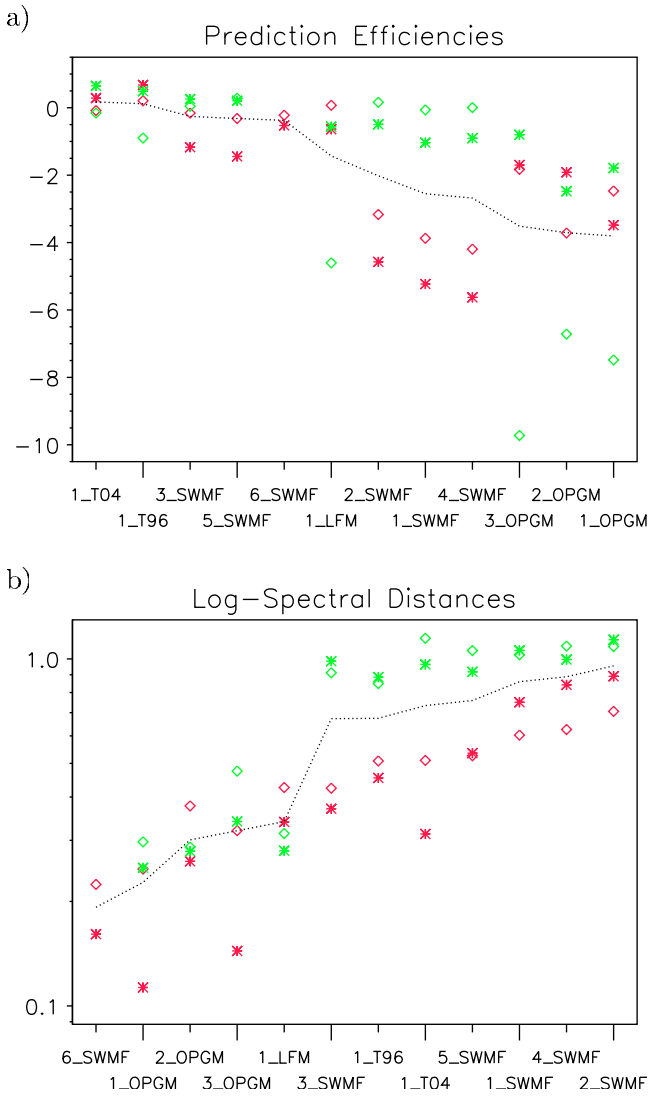

Figure 6. Ranking of model runs for B for weak events (red, event 3, and green, event 4) for (a) prediction efficiency and (b) log-spectral distance.

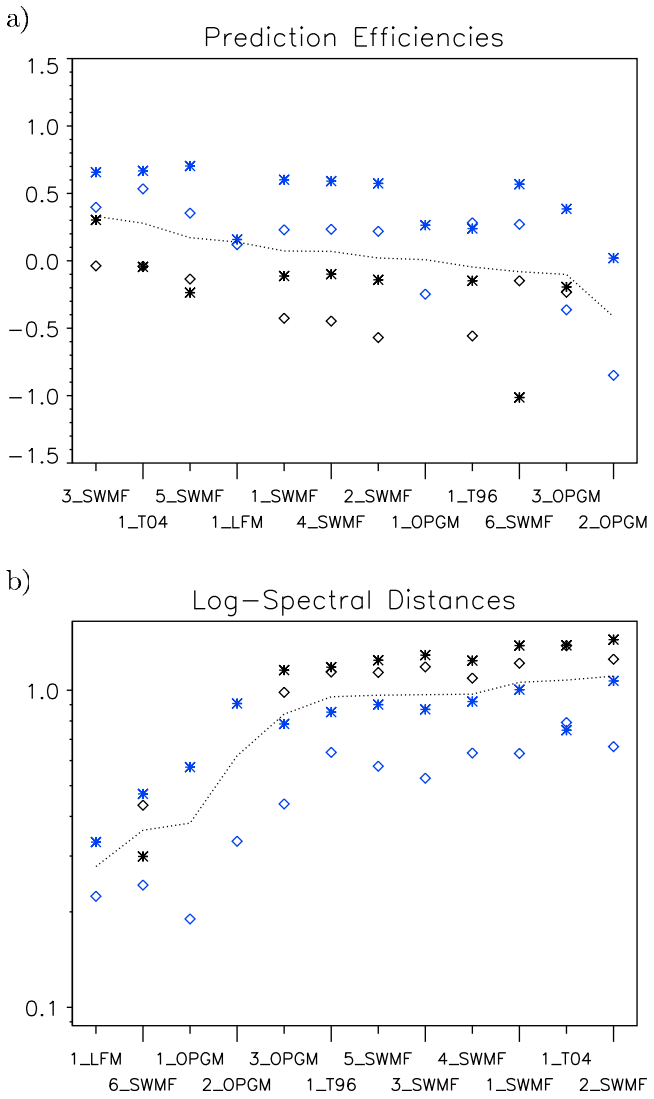

Figure 7. Ranking of model runs for B for strong events for (a) prediction efficiency and (b) log-spectral distance. Note that all models perform worse for event 1 compared to event 2 for both prediction efficiency and log-spectral distance.

and possibly a much-reduced range of values. Figure 10a shows the prediction efficiencies for all events. Indeed, the lowest values (OpenGGCM runs) are not as low as in Figure 5a. Looking at all events in Figure 10a we actually see better scores for T04 and the SWMF model settings for the weaker events (red, event 3; green, event 4) compared to the analysis of B. Also in terms of log-spectral distances in Figure 10b, models seem to do better for weaker events (all red and most of green symbols are below the average for each model setting).

[42] Weaker events show better scores than stronger events (black, event 1; blue, event 2) for better performing models, just opposite of what was found for the prediction efficiencies from the analysis of B. Event 3 (the weakest event of all) does not stand out as consistently producing the worst prediction efficiencies as in Figure 5a. Event 3 yielded the worst prediction efficiencies for many models, which can be attributed to the small denominator σ^2_{obs} in equation (3). With a low variance of observed data, any bias found in modeled results will shift the prediction efficiencies down. In a strongly driven event, a bias will have a much smaller effect. Also, MHD models are strongly driven by the density (N) and magnetic field (B_z) in the solar wind, both of which were in the weak to

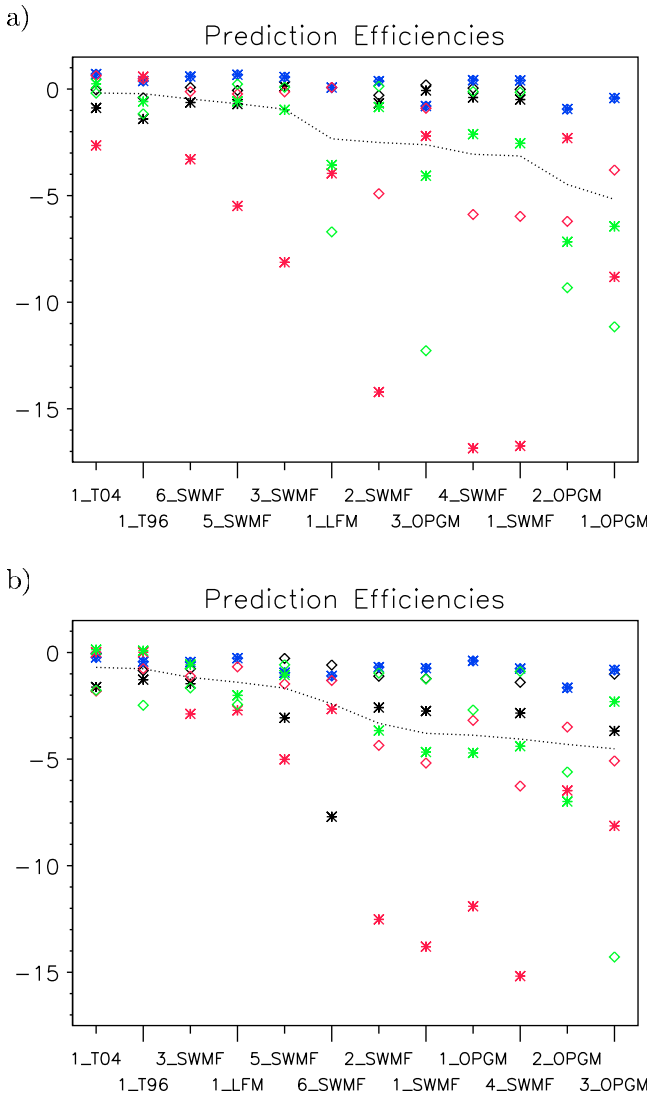

Figure 8. Ranking of model runs for prediction efficiencies for B in (a) nightside (satellite positions $X < 0$) and (b) dayside ($X > 0$) for all events. Stronger events (1 and 2 in black and blue) show higher (positive) scores for the nightside, but average scores taken over all events are very similar between the two regions.

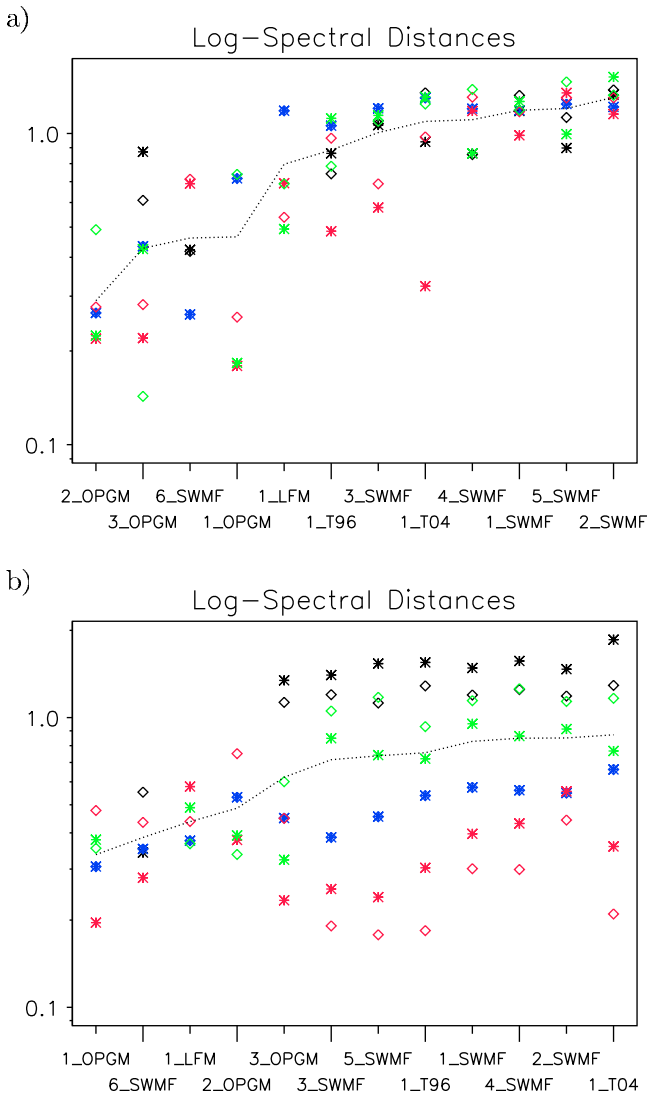

Figure 9. Ranking of model runs in log-spectral distance for **B** in (a) nightside (satellite positions $X < 0$) and (b) dayside ($X > 0$) for all events. In terms of log-spectral distance all models perform better for the dayside than the nightside.

moderate range ($N < 10$ cm^{-3}, $B_z > -5$ nT). This bias effect has been eliminated by looking at τ instead of B.

[43] Figure 11 shows the prediction efficiencies and log-spectral distances for weak events. In comparison to the prediction efficiencies for B shown in Figure 6a, we see in Figure 11a that most scores for τ are much better, except for the scores for OpenGGCM and LFM runs for event 3. Log-spectral distances in Figure 11b are higher than in Figure 6b for 6_SWMF, 1_OPGM, 2_OPGM, 3_OPGM, and 1_LFM runs. For strong events (Figure 12) we find that the scores for all runs and events actually decreased for τ. The range of prediction efficiencies is from −3.2 to +0.2. Comparable prediction efficiencies for B in Figure 7a were between −1.0 and +0.7. The log-spectral distance scores in Figure 10b are generally higher by a factor of 2 for τ than for B in Figure 5.

[44] For SWMF and Tsyganenko model settings for B, events seemed to be ordered by strength (measured by maximum Kp or minimum Dst), where a stronger event

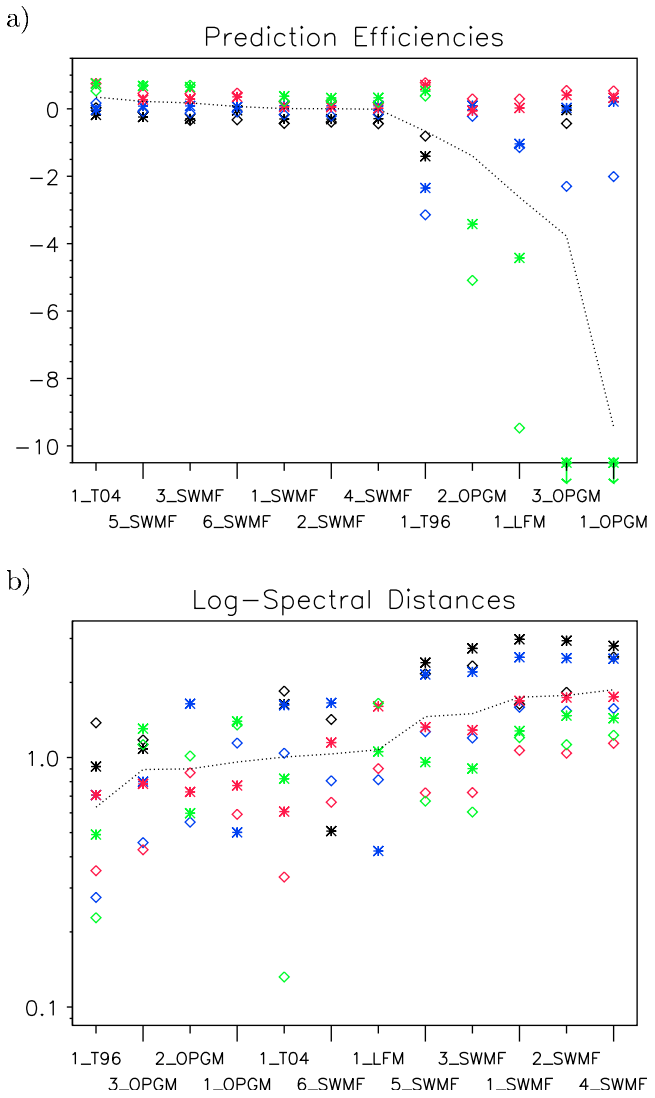

Figure 10. Ranking of model runs for $\tau = \sin(\Theta_{xz})$ according to (a) the prediction efficiency and (b) the log-spectral distance averaged over events and GOES satellites. The off-range (green) data points in Figure 10a for 1_OPGM are at −41.9 (asterisks) and −13.8 (diamonds) and for 3_OPGM are at −13.1 (asterisks) and −15.3 (diamonds). The two strong events (black, event 1; blue, event 2) show lower prediction efficiencies than the weak events for the better performing models (T04, SWMF T96) and are mixed with the weaker events (red, event 3; green, event 4) for the other models (OpenGGCM, LFM).

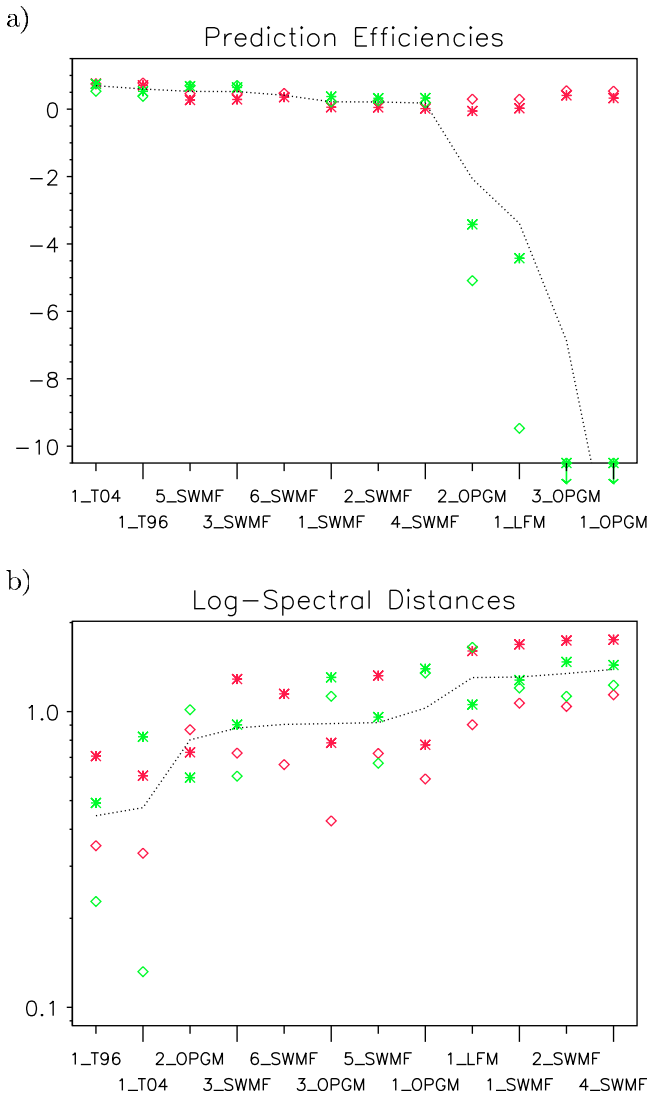

Figure 11. Ranking of model runs for τ for weak events (red, event 3; green, event 4) for (a) prediction efficiency and (b) log-spectral distance. The plot range for Figure 11a is the same as in Figure 10, with the same off-range scores.

results in larger log-spectral distances between observation and modeled specifications. The SWMF and Tsyganenko models score worst (highest log-spectral distances) for event 1, followed by event 3, event 2 and then the weakest event 4. In contrast for τ, event 3 scores worse (higher) than event 4. No obvious ordering could be found between the different events for OpenGGCM and LFM model settings for either B or τ. Whereas individual events score higher or lower, the overall ranking, as an average over all events, remains fairly stable between groups of model settings. In terms of prediction efficiencies, T04 scores best, followed by SWMF, then OpenGGCM. The T96 and LFM runs rank between SWMF and OpenGGCM for B and move into the middle of the SWMF runs for τ. With log-spectral distances, OpenGGCM settings score best as a group, followed by T04 and SWMF. 6_SWMF and LFM and T96 models were different: 6_SWMF and LFM rank higher than OpenGGCM for B and lower for τ and T96 does the opposite (ranking lower for B and best for τ).

6. Discussion

[45] We have presented a first evaluation study of the capability of space weather models to represent the magnetic field conditions at geosynchronous orbit. Multiple simulations of first-principles magnetospheric mod-

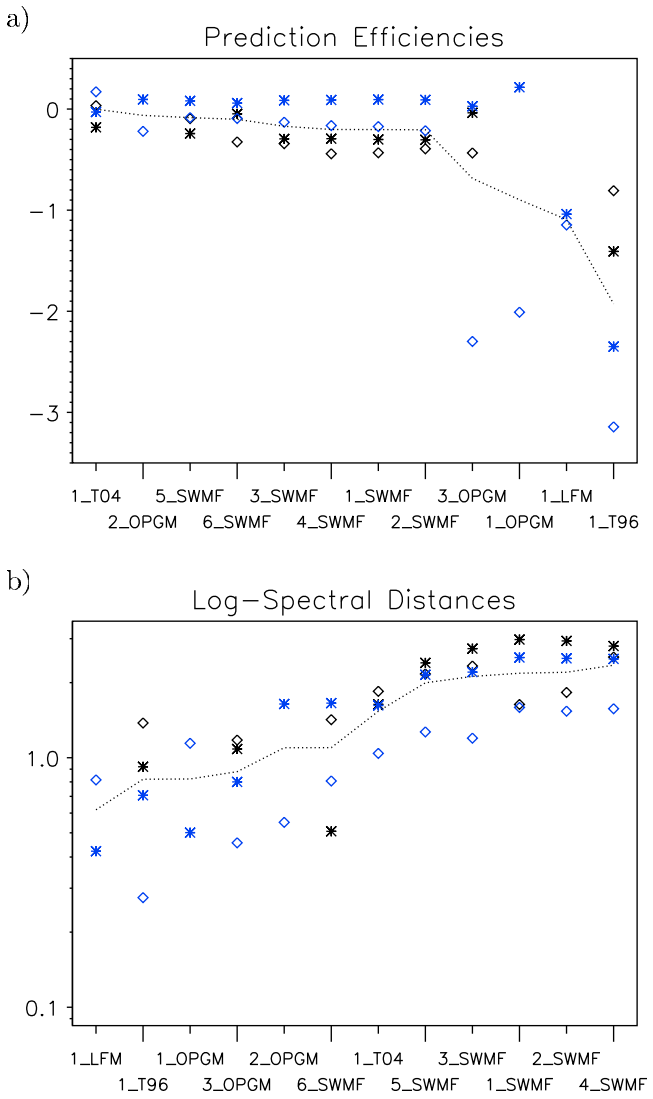

Figure 12. Ranking of model runs for $\tau = \sin(\Theta_{xz})$ for strong events for (a) prediction efficiency and (b) log-spectral distance. Again, the prediction efficiencies for strong events are vastly better than those for weak events. However, the prediction efficiencies are spread into a larger band ($-3.2 < PE < 0.2$) than in Figure 7.

els (six SWMF runs, three OpenGGCM runs, one LFM run) together with outputs of two climatological models (T96 and T04) have been compared with observations obtained at two GOES satellites for four different events (two moderate, two strong storms). We examined two types of skill scores, prediction efficiency and log-spectral distance for two variables, B, the magnetic field strength, and $\tau = \sin(\Theta_{xz})$, a measure of the magnetic elevation angle in the noon-midnight meridian plane. Whereas the evaluations turn out very similar for the two different variables, the ranking of the models vary considerably between the two types of skill scores for each event.

[46] We find that climatological models fare better with weaker events and worse for stronger events. First-principles magnetosphere models did better in strongly driven events. The role of the climatological models T96 and T04 in relation to physics-based models can be described as follows. Both T96 and T04 seem to be optimized to perform best in terms of prediction efficiency as a ratio of the variance between observation and model specification over the variance within the observation data. As expected, T96 performs best for moderate storm events. For stronger storm events, the input exceeds the range of validity and T96 becomes less useful. T04 is a major improvement over T96 as it is designed to have an extended validity and we see that this model still outperforms first-principles models in terms of prediction efficiency for all events.

[47] In terms of log-spectral distance which measures the coverage of a range of temporal scales between 1 min and 1 h, the physics-based models perform better than the climatological models. Several (but not all) settings of physics-based models exceed the performance of both climatological Tsyganenko models (except for T96 which performed best for variable τ).

[48] Physics-based models perform as well or better for large events than moderate or weak events. Especially, prediction efficiencies of B increased for strong storm events. Since physics-based models can compute a magnetosphere solution from any solar wind input, the limits of validity are, in principle, larger than those of climatological models which may lack data coverage for events with extreme solar wind parameters. On the other hand, physics-based models have a large array of internal (numerical) parameters, most important of which is the grid resolution, location of the inner magnetosphere boundary and the presence of an inner magnetospheric kinetic physics module. SWMF, for instance, benefits from the addition of the Rice Convection Model (RCM) for the inner magnetosphere, more than it benefits from an increase in resolution (3 million cells at a finest resolution of $0.125R_E$ in the inner magnetosphere and near-Earth plasma sheet, compared to 2 million at $0.25R_E$ resolution). Due to the design of the OpenGGCM model (no adaptive grid) doubling the number of grid cells could only be accompanied with a slight increase of resolution (from 0.3 to $0.25R_E$) with no discernible change in performance. SWMF, in general, features more smoother solutions than OpenGGCM or LFM, resulting in lower performance in terms of the log-spectral distance.

[49] In addition to looking at two quantities that were derived from the observations and models, we investigated the performance of models with respect to two different regions in the magnetosphere (satellite location on the nightside and dayside). Although there are slight changes in the ranking of individual models, the overall rankings are very similar to the original rankings for the whole events. Even though the relative ranking did not change greatly between the regions, the numerical values for the individual scores (especially the prediction efficiency) can change considerably when looking only at the nightside or dayside, as opposed to both regions combined in one calculation. This emphasizes that one has to be careful about comparing numerical scores out of context. The same time period during an event has to be used in a calculation of a score in future challenge campaigns, with modified and improved models.

[50] We recognize that examining a small set of parameters (magnetic field magnitude and orientation in planes parallel to the noon-midnight plane) addresses only a small region in the magnetosphere and a small aspect of magnetospheric modeling. We, however, think that an assessment of the fidelity of magnetic field specification at geosynchronous orbit sufficiently addresses the need for space weather forecasting in a region used by a large number of operational satellites. A good magnetic field forecast is essential for other space weather models such as those that describe the radiation belts and ring currents or the propagation and impact of solar and galactic energetic particles.

[51] Magnetospheric MHD models essentially solve the same set of (ideal MHD) equations with a coupling to a two-dimensional ionospheric electrodynamic solver. The differences in the numerical schemes and resolution are reflected in the different magnetic field prediction efficiencies and spectral responses. Models with low diffusion, such as LFM, OpenGGCM and SWMF with the Sokolov solver, perform better at the log-spectral distance analysis. It is found that inclusion of an RCM module in SWMF simulations leads to an increase of the plasma pressure in the inner magnetosphere and thus an increase of the volume with closed magnetic field lines.

[52] This study is a first step in a planned series that invites modelers to improve their models and submit results of new simulation runs. Over time the performance of each type of simulation model can be reevaluated and the improvement in performance will be seen over the years as models mature. Operational agencies such as NOAA (Space Weather Prediction Center) and Air Force (Air Force Weather Agency), among others, are interested to see those evaluations and use them to make decisions about incorporating physics-based models of the global magnetosphere into space weather operations.

[53] Depending on the type of application, different models may be better suited to provide specifications. It is important to mention that for almost all models in this

study, the scores obtained for the studied events can vary almost as much as the average performances vary between the models. Although we have seen that groups of models consistently rank high and others low for each type of skill score, the statistical significance of those difference is low and further studies have to be performed (e.g., include more events beyond the four events studied) to confirm those rankings.

[54] **Acknowledgments.** Some model simulations were obtained at the Community Coordinated Modeling Center. GOES 1 min magnetic data were obtained from the NASA Space Physics Data Facility's CDAWeb interface. Satellite orbit data were obtained from the NASA Satellite Situation Center's SSCWeb interface. The authors thank the two referees for their comments which greatly improved this paper.

References

Birn, J., et al. (2001), Geospace Environmental Modeling (GEM) Magnetic Reconnection Challenge, *J. Geophys. Res.*, *106*(A3), 3715–3719.

Lyon, J. G., J. A. Fedder, and C. M. Mobarry (2004), The Lyon-Fedder-Mobarry (LFM) global MHD magnetospheric simulation code, *J. Atmos. Sol. Terr. Phys.*, *66*, 1333–1350.

Lyons, L. R. (1998), The Geospace Modeling Program Grand Challenge, *J. Geophys. Res.*, *103*(A7), 14,781–14,785.

Merkin, V. G., M. J. Owens, H. E. Spence, W. J. Hughes, and J. M. Quinn (2007), Predicting magnetospheric dynamics with a coupled Sun-to-Earth model: Challenges and first results, *Space Weather*, *5*, S12001, doi:10.1029/2007SW000335.

Powell, K. G., P. L. Roe, T. J. Linde, T. I. Gombosi, and D. L. De Zeeuw (1999), A solution-adaptive upwind scheme for ideal magnetohydrodynamics, *J. Comput. Phys.*, *154*, 284–309.

Pulkkinen, A., L. Rastätter, M. Kuznetsova, M. Hesse, A. Ridley, J. Raeder, H. J. Singer, and A. Chulaki (2010), Systematic evaluation of ground and geostationary magnetic field predictions generated by global magnetohydrodynamic models, *J. Geophys. Res.*, *115*, A03206, doi:10.1029/2009JA014537.

Pulkkinen, A., et al. (2011), Geospace Environment Modeling 2008–2009 Challenge: Ground magnetic field perturbations, *Space Weather*, *9*, S02004, doi:10.1029/2010SW000600.

Raeder, J., and N. Maynard (2001), Foreword, *J. Geophys. Res.*, *106*(A1), 345–348.

Raeder, J., R. L. McPherron, L. A. Frank, S. Kokubun, G. Lu, T. Mukai, W. R. Paterson, J. B. Sigwarth, H. J. Singer, and J. A. Slavin (2001), Global simulation of the geospace environment modeling substorm challenge event, *J. Geophys. Res.*, *106*(A1), 381–395.

Ridley, A. J., K. C. Hansen, G. Tóth, D. L. De Zeeuw, T. I. Gombosi, and K. G. Powell (2002), University of Michigan MHD results of the Geospace Global Circulation Model metrics challenge, *J. Geophys. Res.*, *107*(A10), 1290, doi:10.1029/2001JA000253.

Ridley, A. J., T. I. Gombosi, and D. L. De Zeeuw (2004), Ionospheric control of the magnetospheric configuration: Conductance, *Ann. Geophys.*, *22*, 567–584.

Ridley, A. J., T. I. Gombosi, I. V. Sokolov, G. Tóth, and D. T. Welling (2010), Numerical considerations in simulating the global magnetosphere, *Ann. Geophys.*, *28*, 1589–1614, doi:10.5194/angeo-28-1589-2010.

Rufenach, C. L., R. F. Martin, and H. H. Sauer (1989), A study of geosynchronous magnetopause crossings, *J. Geophys. Res.*, *94*(A11), 15,125–15,134.

Sazykin, S., R. A. Wolf, R. W. Spiro, T. I. Gombosi, D. L. De Zeeuw, and M. F. Thomsen (2002), Interchange instability in the inner magnetosphere associated with geosynchronous particle flux decreases, *Geophys. Res. Lett.*, *29*(10), 1448, doi:10.1029/2001GL014416.

Skoug, R. M., J. T. Gosling, J. T. Steinberg, D. J. McComas, C. W. Smith, N. F. Ness, Q. Hu, and L. F. Burlaga (2004), Extremely high speed solar wind: 29–30 October 2003, *J. Geophys. Res.*, *109*, A09102, doi:10.1029/2004JA010494.

Sokolov, I. V., E. V. Timofeev, J. Sakai, and K. Takayama (1999), On shock-capturing schemes using artificial wind, *Shock Waves*, *9*, 423–426.

Tóth, G., et al. (2005), Space Weather Modeling Framework: A new tool for the space science community, *J. Geophys. Res.*, *110*, A12226, doi:10.1029/2005JA011126.

Tóth, G., D. L. De Zeeuw, T. I. Gombosi, and K. G. Powell (2006), A parallel explicit/implicit time stepping scheme on block-adaptive grids, *J. Comput. Phys.*, *217*, 722–758.

Tsyganenko, N. A. (1995), Modeling the Earth's magnetospheric magnetic field confined within a realistic magnetopause, *J. Geophys. Res.*, *100*(A4), 5599–5612.

Tsyganenko, N. A. (1996), Effects of the solar wind conditions on the global magnetospheric configuration as deduced from data-based field models, in *Third International Conference on Substorms (ICS-3), Versailles, France, 12–17 May 1996*, edited by E. J. Rolfe and B. Kaldeich, *Eur. Space Agency Spec. Publ., ESA SP 389*, 181–185.

Tsyganenko, N. A., and M. I. Sitnov (2005), Modeling the dynamics of the inner magnetosphere during strong geomagnetic storms, *J. Geophys. Res.*, *110*, A03208, doi:10.1029/2004JA010798.

Welling, D. T., and A. J. Ridley (2010), Validation of SWMF magnetic field and plasma, *Space Weather*, *8*, S03002, doi:10.1029/2009SW000494.

Wiltberger, M., T. I. Pulkkinen, J. G. Lyon, and C. C. Goodrich (2000), MHD simulation of the magnetotail during the December 10, 1996, substorm, *J. Geophys. Res.*, *105*(A12), 27,649–27,663.

Wiltberger, M., W. Wang, A. G. Burns, S. C. Solomon, J. G. Lyon, and C. C. Goodrich (2004), Initial results from the coupled magnetosphere ionosphere thermosphere model: Magnetospheric and ionospheric responses, *J. Atmos. Sol. Terr. Phys.*, *66*, 1411–1423.

M. Hesse, M. M. Kuznetsova, A. Pulkkinen, and L. Rastätter, Community Coordinated Modeling Center, Space Weather Laboratory, NASA Goddard Space Flight Center, Code 674, Greenbelt, MD 20770, USA. (michael.hesse@nasa.gov; maria.m.kuznetsova@nasa.gov; antti.a.pulkkinen@nasa.gov; lutz.rastaetter@nasa.gov)

A. Ridley, Department of Atmospheric, Oceanic, and Space Sciences, 1416 Space Research Bldg., University of Michigan, Ann Arbor, MI 48109-2143, USA. (ridley@umich.edu)

H. J. Singer, Space Weather Prediction Center, National Centers for Environmental Prediction, NOAA National Weather Service, 325 Broadway, Boulder, CO 80305, USA. (howard.singer@noaa.gov)

A. Vapirev, Space Science Center, University of New Hampshire, 254 Morse Hall, 8 College Rd., Durham, NH 03824, USA. (alexander.vapirev@unh.edu)

M. Wiltberger, High Altitude Observatory, National Center for Atmospheric Research, Boulder, CO 80307-3000, USA. (wiltbemj@ucar.edu)