

On the performance of global magnetohydrodynamic models in the Earth's magnetosphere

I. Honkonen,^{1,2} L. Rastätter,³ A. Grocott,⁴ A. Pulkkinen,^{3,5} M. Palmroth,¹ J. Raeder,⁶ A. J. Ridley,⁷ and M. Wiltberger⁸

Received 13 February 2013; revised 25 April 2013; accepted 28 April 2013; published 30 May 2013.

[1] We study the performance of four magnetohydrodynamic models (BATS-R-US, GUMICS, LFM, OpenGGCM) in the Earth's magnetosphere. Using the Community Coordinated Modeling Center's Run-on-Request system, we compare model predictions with magnetic field measurements of the Cluster, Geotail and Wind spacecraft during a multiple substorm event. We also compare model cross polar cap potential results to those obtained from the Super Dual Auroral Radar Network (SuperDARN) and the model magnetopause standoff distances to an empirical magnetopause model. The correlation coefficient (CC) and prediction efficiency (PE) metrics are used to objectively evaluate model performance quantitatively. For all four models, the best performance outside geosynchronous orbit is found on the dayside. Generally, the performance of models decreases steadily downstream from the Earth. On the dayside most CCs are above 0.5 with CCs for B_x and B_z close to 0.9 for three out of four models. In the magnetotail at a distance of about -130 Earth radii from Earth, the prediction efficiency of all models is below that of using an average value for the prediction with the exception of B_z . B_x is most often best predicted and correlated both on the dayside and the nightside close to the Earth whereas in the far tail the CC and PE for B_z are substantially higher than other components in all models. We also find that increasing the resolution or coupling an additional physics module does not automatically increase the model performance in the magnetosphere.

Citation: Honkonen, I., L. Rastätter, A. Grocott, A. Pulkkinen, M. Palmroth, J. Raeder, A. J. Ridley, and M. Wiltberger (2013), On the performance of global magnetohydrodynamic models in the Earth's magnetosphere, *Space Weather*, 11, 313–326, doi:10.1002/swe.20055.

¹Finnish Meteorological Institute, Helsinki, Finland.

²Department of Physics, University of Helsinki, Helsinki, Finland.

³Community Coordinated Modeling Center, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

⁴Department of Physics and Astronomy, University of Leicester, Leicester, UK.

⁵Institute for Astrophysics and Computational Sciences, Catholic University, Washington, DC, USA.

⁶Space Science Center and Physics Department, University of New Hampshire, Durham, New Hampshire, USA.

⁷Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, Michigan, USA.

⁸High Altitude Observatory, National Center for Atmospheric Research, Boulder, Colorado, USA.

Corresponding author: I. Honkonen, Earth Observation, Finnish Meteorological Institute, PL 503, 00101, Helsinki, Finland. (ilja.honkonen@fmi.fi)

©2013. American Geophysical Union. All Rights Reserved.
1542-7390/13/10.1002/swe.20055

1. Introduction

[2] The growing interest in space weather forecasting from both government and industry is also increasing the need for space weather model development. The increasing amount of infrastructure and people that can be affected by severe space weather events demand reliable forecasting of those events and their effects both in space and on the ground. This in turn requires systematic testing, verification and validation of space weather models and the objective evaluation of their suitability for a particular purpose. The international need for model improvement and validation was highlighted, for example, in the recent European Commission's Space Weather Awareness Dialogue [Krausmann and Bothmer, 2012].

[3] A good overview of the process of verifying and validating a space plasma model is given by Ledvina *et al.* [2008]. The verification of a code consists of, for example, comparing the results to an analytic solution (e.g., using

the method of manufactured solutions employed recently by *Welling et al.* [2012] for verifying SpacePy); monitoring conserved quantities, symmetries, and other predictable outcomes; or comparing results to those from other codes. Verification must happen before validation in order to make sure that the equations chosen for modeling the system are being solved correctly, and the results should be published especially in the case of a new model or a scheme. In this regard global magnetohydrodynamic (MHD) models leave something to be desired as only the BATS-R-US (Block Adaptive Tree Solar-wind Roe Upwind Scheme) code has been verified against analytic or semi-analytic results in a publication [*Powell et al.*, 1999]. Results for the ubiquitous shock tube test (see *Ryu and Jones* [1995] for a plethora of examples) have not been presented for the MHD solver(s) of any global MHD model even though such tests have been conducted for all models used here. For example, the MHD solvers of GUMICS were used in several one, two, and three-dimensional tests by *Honkonen et al.* [2013], but in addition to the parallel scalability results of all tests, only the result of a three-dimensional blast wave test with adaptive mesh refinement (AMR) was shown. The validation of a code can consist of controlled experiments designed to investigate a physical process, experiments specifically designed to validate codes or passive observations of physical events. Global MHD models have mostly been validated with many qualitative comparisons against observations by various spacecraft and, to our knowledge, no systematic quantitative comparisons have been conducted until recently.

[4] The Geospace Environment Modeling (GEM) 2008–2009 challenge represents the largest effort to date to validate global MHD models against observations using various objective metrics. *Pulkkinen et al.* [2011] quantified the performance of three global MHD models and two empirical models in reproducing observations of ground magnetic field during four geospace storm events. The models were compared against observations from 12 observatories located between 43.5° and 74° of geomagnetic latitude. Five different metrics, each applicable to different situations, were used to evaluate the model performance: root-mean-square difference, prediction efficiency, log-spectral distance, utility, and ratio of maximum amplitudes. *Rastätter et al.* [2011] used the same events and global MHD models as *Pulkkinen et al.* [2011] along with two empirical magnetospheric models to quantify model performance at the geosynchronous orbit with respect to observations of magnetic field strength and elevation. The models were compared to observations from two NOAA Geosynchronous Operational Environmental Satellites (GOES) in each event using the prediction efficiency and log-spectral distance metrics. The most recent GEM 2008–2009 challenge comparison [*Rastätter et al.*, 2013] quantified the ability of different physics-based and statistical models to reproduce the Dst geomagnetic activity index (<http://wdc.kugi.kyoto-u.ac.jp/dstdir>).

[5] Magnetospheric substorms are an essential part of the dynamics of near-Earth space [*McPherron*, 1991]: During the substorm growth phase, magnetic flux accumulates in the tail lobes due to dayside reconnection and is eventually released by rapid reconnection in the near-Earth magnetotail, which also starts the substorm expansion phase. The cross-tail current is diverted from the region of reconnection and flows along magnetic field lines to the midnight ionosphere where it flows westward for a (relatively) short distance enhancing locally the westward electrojet before returning to the tail. Magnetospheric substorms and hence the dynamics of the near and far tail magnetosphere are also important from the point of view of space weather and can have a significant effect on technological systems even at ground level. For example, the largest geomagnetically induced currents (GIC) occur with highest probability during the substorm expansion phase about 5 min after the expansion onset below the corrected geomagnetic (CGM) latitude of 72° [*Viljanen et al.*, 2006].

[6] In this work, the performance of four global MHD models is systematically evaluated in the Earth's magnetosphere and in the ionosphere during an event with multiple substorms on 18 Feb 2004. The global MHD models BATS-R-US (Block Adaptive Tree Solar-wind Roe Upwind Scheme), GUMICS-4 (Grand Unified Magnetosphere-Ionosphere Coupling Simulation), LFM (Lyon-Fedder-Mobarry), and OpenGGCM (Open General Geospace Circulation Model) are given identical solar wind input, and the results are compared to the magnetic field measurements of Cluster 1 [*Balogh et al.*, 1997] within the magnetosheath, Geotail [*Kokubun et al.*, 1994] in the near tail, Wind [*Lepping et al.*, 1995] in the far tail, and the cross polar cap potential (CPCP) obtained from SuperDARN [*Chisham et al.*, 2007]. The models' magnetopause standoff distances are also compared to the empirical magnetopause model of *Lin et al.* [2010]. All simulations are carried out through NASA's Community Coordinated Modeling Center (CCMC) Run-on-Request system (<http://ccmc.gsfc.nasa.gov>), and the settings used for the models are as close to each other as reasonably possible. The results presented here are also available through CCMC. Similarly to *Pulkkinen et al.* [2011] and *Rastätter et al.* [2011], the detailed scientific analysis of the effect of various model parameters on the quality of model results in the magnetosphere is left for future work. As *Pulkkinen et al.* [2011] and *Rastätter et al.* [2011] studied the model performance at ground level and at geosynchronous orbit, this study is a natural next step to these investigations by validating the code performance in the near and far tail during dynamical events. In section 2, we describe the models, the features, and parameters which were used in this work and the event that was simulated. In section 4, we present the model results with the corresponding measurements, and in section 5, we compare them using the correlation coefficient (CC) and prediction efficiency (PE) metrics. We discuss the results and analysis in section 6 and draw our conclusions in section 7.

Table 1. Summary of Features and Settings of Global MHD Models Used in This Study, See the Text for Details

	BATS-R-US	GUMICS-4	LFM	OpenGGCM
MHD equations	ideal, conservative, $B_0 + B_1$	ideal, conservative, $B_0 + B_1$	ideal, semi-conservative, $B_0 + B_1$	semi-conservative with resistivity
Solver notes	eight-wave approximate Riemann	mostly Roe, subcycling, $\nabla \cdot B$ cleaning	total variation diminishing (TVD), constrained transport (CT)	TVD, CT
Order of MHD discretization: spatial / temporal	2 / 2	1 / 1	8 / 2	4 / 2
MHD grid	Cartesian, static, block-refined	Cartesian, dynamic, cell-refined	distorted spherical, static, not refined	stretched Cartesian, static, not refined
Dipole tilt updated with time	yes	no	yes ^a	no
Coordinate system of magnetosphere	GSM	GSE	SM	GSE

^aThe dipole orientation is fixed in SM coordinates, but solar wind and solar EUV conditions are adjusted with time.

2. Global MHD Model Features and Settings

[7] The features and settings of global MHD models used in this study are presented in Table 1. We emphasize that some of the models support a wide range of features and different settings, but we have listed only the ones used in this study. All models are executed through the CCMC Run-on-Request system and receive as input the solar wind data measured by the Advanced Composition Explorer (ACE) satellite [Stone *et al.*, 1998] located at GSE (221, -22, 9) R_E (Earth radii) during the simulated event provided by CCMC. The options and features used in all models are as close to each other as reasonably possible.

[8] While all the models solve the MHD equations in the magnetosphere and the same electrostatic potential equation in the ionosphere, there are some differences. Both BATS-R-US and GUMICS solve the ideal (i.e., inviscid and perfectly conducting), conservative, non-relativistic MHD equations [Powell *et al.*, 1999; Janhunen *et al.*, 2012], while LFM and OpenGGCM solve the MHD equations in a semi-conservative form where the total energy is replaced with the fluid energy [Lyon *et al.*, 2004; Raeder *et al.*, 2008]. In OpenGGCM, a resistive term is also included in the equation for the electric field [Raeder *et al.*, 2008]. BATS-R-US solves the MHD equations using a second-order eight-wave approximate Riemann solver that maintains zero divergence of the magnetic field to truncation error. GUMICS primarily uses a first-order seven-wave approximate Riemann solver and is the only model to periodically remove the divergence of magnetic field with the projection method of Brackbill and Barnes [1980]. Both LFM and OpenGGCM use a constrained transport method for advecting the magnetic field which preserves the divergence of magnetic field to roundoff error. BATS-R-US and GUMICS use an adapted Cartesian mesh for the magnetosphere. In BATS-R-US, the grid is adapted at the start of the simulation in blocks of 6^3 cells and is static afterwards, while in GUMICS, the grid is adapted during the simulation on a cell-by-cell basis

based on local gradients of several plasma quantities and geometric considerations. LFM uses a distorted spherical grid and OpenGGCM uses a stretched Cartesian grid and neither uses AMR. BATS-R-US, GUMICS and LFM separate the magnetic field into perturbed and static background components [see Tanaka 1994]. The number of cells in the magnetospheric grid are about 800 k in BATS-R-US, 400 k in GUMICS, 330 k in LFM, and 3.6 M in OpenGGCM.

[9] The inner boundary of the magnetosphere in the models is between 2 and 4 R_E . The ionosphere and magnetosphere are coupled through field aligned currents (FAC) and electric potentials mapped between the ionosphere and the inner boundary of the magnetosphere along the Earth's dipole magnetic field. Field aligned currents are obtained from currents computed from the magnetic field in the inner magnetosphere. A two-dimensional electrostatic solver is used in the ionosphere to solve the ionospheric potential from FACs using the current continuity equation. The solved electric potential is used to set plasma flow in the inner magnetosphere. Merkin and Lyon [2010] provides more details on the LFM potential solver. OpenGGCM also includes a three-dimensional dynamical model of the thermosphere which adds, for example, the effect of the neutral wind dynamo to the ionospheric electric potential solution.

3. Event Description and Data

[10] Magnetospheric substorms are an essential part of the dynamics of near-Earth space [McPherron, 1991], and hence, an event with multiple substorms was selected to assess the performance of global models in the Earth's magnetosphere. The solar wind at the ACE satellite during the event of 18 Feb 2004 along with AE Dst indices is shown in Figure 1. The delay from ACE to the magnetopause for the event was calculated by Honkoniemi *et al.* [2011] to be 46 min. During the event, the solar wind density fluctuated between 1 and 3 cm^{-3} with large jumps

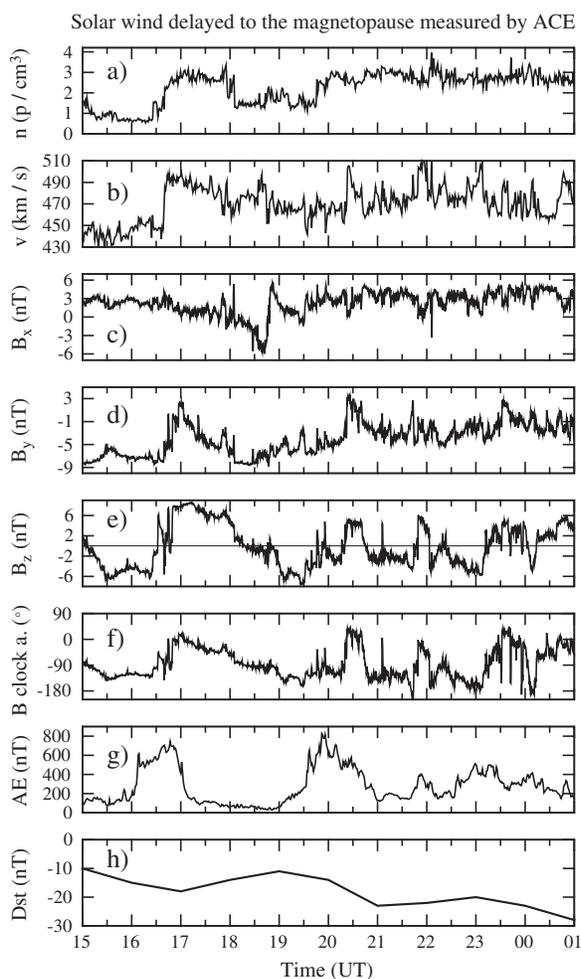


Figure 1. (a) Density, (b) velocity, (c) B_x , (d) B_y , (e) B_z , and (f) clock angle of the delayed (46 min) solar wind measured by ACE on 18 Feb 2004 at $x_{GSE} = 221 R_E$, (g) the provisional AE index and (h) final Dst index from Kyoto index service. Adapted from *Honkonen et al.* [2011].

recorded at 16:40 and 19:45 UT. The interplanetary magnetic field (IMF) z component changes sign more than a dozen times, varying between -8 and 8 nT. This leads to modest driving of the magnetosphere-ionosphere system as indicated by the AE index being over 600 for several hours during the event. Solar wind velocity is slightly above the average staying between 430 and 490 km/s.

[11] Undelayed solar wind is used as input for all models and, consequently, the delay from ACE to the solar wind boundary of the models must be taken into account separately. We assume a constant solar wind speed when calculating the delay for each model. In BATS-R-US the solar wind boundary is located at $33 R_E$ which translates to a delay of 2505 s. In GUMICS, LFM and OpenGGCM the solar wind boundaries are at 32, 30 and $60 R_E$, which translate to delays of 2518, 2545 and 2145 s, respectively.

These delays were used for all model results when comparing against observations.

[12] The simulation results reported here are available through the CCMC Web site as run numbers 102709_1, 103009_1, 011110_1, and 020410_1 prefixed with “Ilja_Honkonen_.” The data from all simulations used in this study is saved to disk every 5 min, hence, all the satellite data is averaged using a 5 min sliding window. Figure 2 shows the trajectories of Cluster 1, Geotail and Wind during the simulated event with rectangles marking their location at the start of the event (15:00 UT). The distances of Cluster, Geotail, and Wind are about 9, 26, and $133 R_E$ from Earth, respectively. Cluster is flying towards dusk side ecliptic plane, Geotail is advancing outward from the dawn side flank, and Wind is almost stationary in the far tail.

4. Results

[13] Figure 3 shows the magnetic field components from simulations and Cluster 1 as a function of time, CC, and PE scores calculated from that data in section 5, scatter plots of each model B component versus the observation at the same instant of time, the coordinate of Cluster 1, and the region in which it is located. Based on Cluster 1 ion energy spectrogram data (not shown), Cluster was in the magnetosheath (labeled msheath in the Figure) with short excursions into the magnetosphere (labeled ms) until about 23:00 UT after which Cluster moved into the solar wind (labeled sw). Before 19:00 UT, there is a large difference between the models’ B_x prediction at Cluster 1 but afterwards models and Cluster show a fairly constant B_x . B_x has two noticeable depressions at 20:30 and 22:00 UT of which the first is captured by BATS-R-US, GUMICS, and LFM and the second by OpenGGCM and perhaps by BATS-R-US. B_y shows several large changes between 17:00 and 23:00 UT. The largest change in B_y around 17:00 UT is captured by BATS-R-US, GUMICS, and LFM. The increase at 20:15 UT is reproduced best by BATS-R-US and GUMICS while in OpenGGCM and LFM, the increase

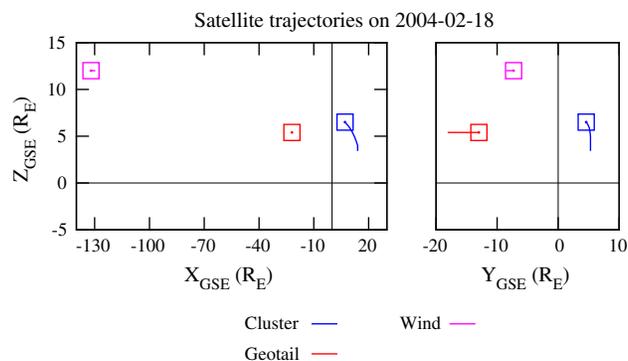


Figure 2. Trajectories of Cluster 1, Geotail, and Wind during the simulated event with rectangles marking their location at the start of the event (15:00 UT).

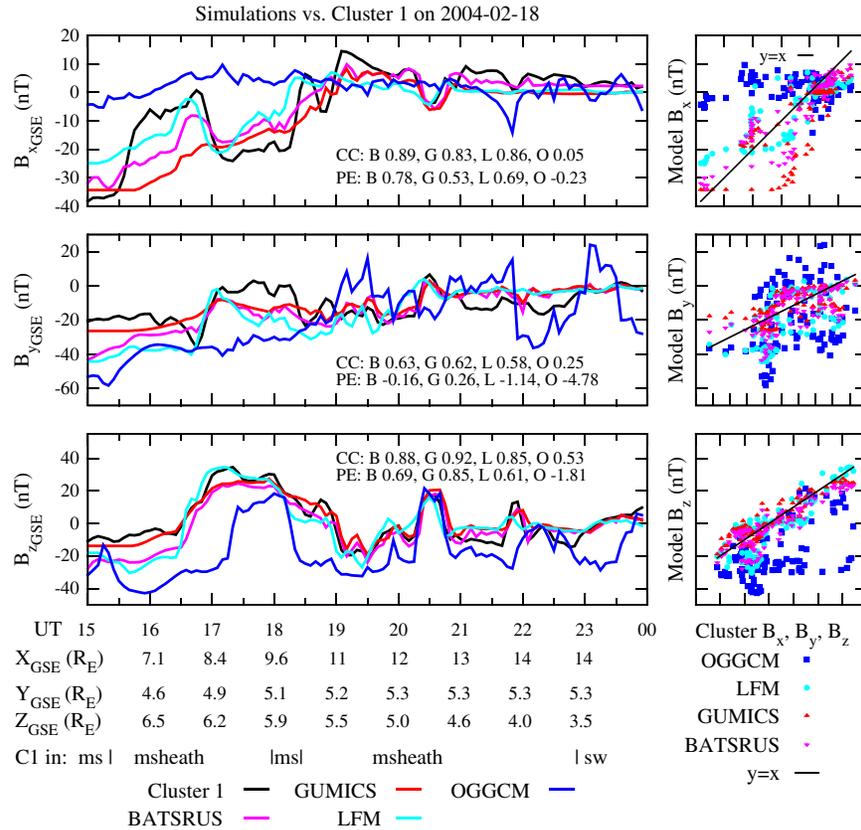


Figure 3. Five minute magnetic field data from Cluster 1 obtained by averaging spin average data over a 5 min window and magnetic field data at the same location from four global MHD simulations as function of time and scatter plots of each model B component versus the observation at the same instant of time. Correlation coefficients and prediction efficiencies between model results and observations are also shown, see section 5 for details. Location and region of Cluster 1 are shown at the bottom, see the text for an explanation of region codes.

starts some 20 min earlier. B_z shows three large increases and subsequent decreases starting around 16:30, 20:15, and 21:30 UT. The major features of B_z are captured by all models. After 17:00 UT, BATSRUS and GUMICS are in good agreement with each other and along with LFM are close to Cluster 1 observations. OpenGGCM reproduces Cluster 1 observations reasonably well but with an offset of about -10 nT and a 1 to 2 h shorter first enhancement in B_z at 17:00 UT. At 20:30 UT, all models are in good agreement with Cluster 1. The temporary spread in modeled results at 18:30 UT stands out. For all components of B , the largest differences between models occur at the beginning of the event before about 17:00 UT.

[14] Figure 4 shows the magnetic field data from simulations and Geotail as a function of time. Based on density and ion temperature data, Geotail is mostly in the plasma sheet (labeled ps in the Figure) or plasma sheet boundary layer (labeled psbl) before 21:00 UT and in the lobe afterwards [Aikio *et al.*, 2013]. B_x has large changes around 16:00 and 20:00 UT of which only the changes before 17:30 UT

seem to be reproduced by the models. All models predict the sharp decrease in B_x at Geotail around 17:00 UT. Geotail shows two depressions in B_x around 20:00 and 23:00 UT, but on the other hand, B_x stays more or less constant in BATSRUS and LFM. Furthermore in GUMICS and OpenGGCM, contrary to Geotail, B_x clearly increases around 20:00 UT and also seems to increase, on average, around 23:00 UT. B_y does not show large features at Geotail except for noticeable increases at 16:00 and 17:00 UT. The former one seems to be captured by BATSRUS and OpenGGCM while the latter is captured by LFM and, with a small delay, by BATSRUS and GUMICS. The steady decrease of B_y between 17:00 and 19:30 UT is reproduced best by LFM with BATSRUS and GUMICS also showing a similar feature. Geotail B_z has two large increases starting at 17:00 and 20:00 UT, which last for several hours. Between about 17:00 and 21:00 UT, all the models agree quite well with each other and, with an added offset of about 3 nT, also with Geotail. Interestingly, B_z in BATSRUS, GUMICS, and LFM at Geotail is quite similar

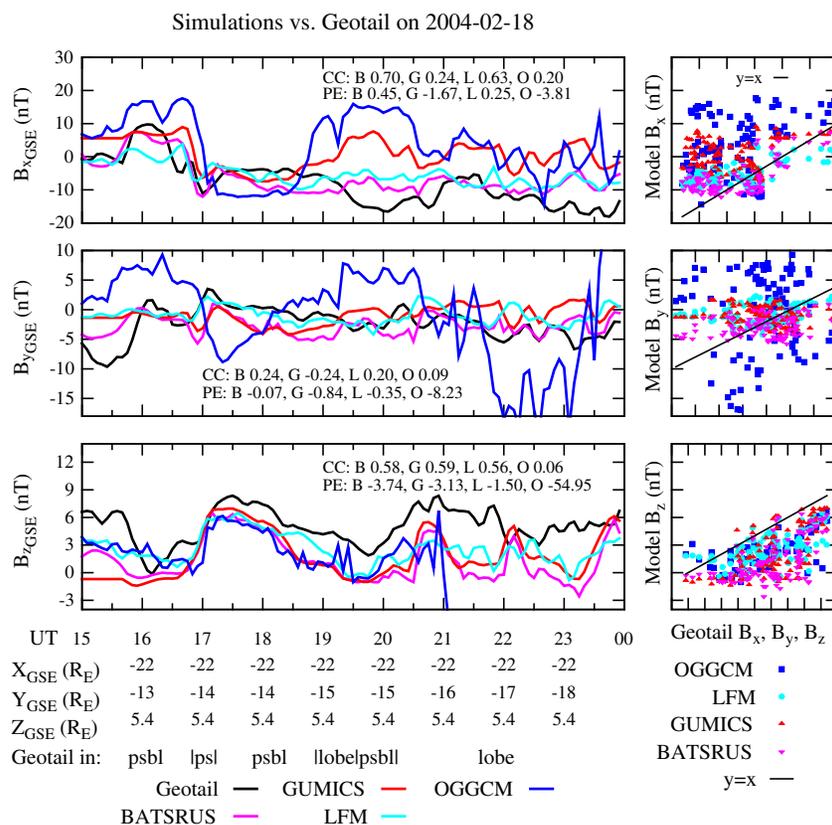


Figure 4. Magnetic field data from Geotail and simulations shown in the same format as in Figure 3.

to the simulated B_z at Cluster 1, but with a delay of 15 to 30 min. This is not the case for observations where Geotail shows one large enhancement between 20:00 and 22:00 UT while Cluster 1 shows two separate smaller enhancements at 20:30 and 22:00 UT. The disagreement between Geotail B_z and the models is largest before 16:00, at 21:30 and at 23:00 UT.

[15] Figure 5 shows the magnetic field data from simulations and Wind as a function of time. During the event, Wind travels between the northern or southern lobes via the neutral sheet (labeled nsheet in the Figure). On average B_x seems to decrease steadily between 16:00 and 19:00 UT after which it starts to increase. Several depressions of B_x are overlaid on top of the average behavior, the largest ones being at about 17:30, 19:30, 21:00, and 22:30 UT. The only features that seem to be reproduced consistently by all models are the depressions of B_x at 17:30 and 21:00 UT. OpenGGCM shows quite good agreement with Wind between 19:30 and 21:00 UT. Wind shows large increases in B_y starting at 17:00, 18:30, and 21:00 UT, and all models reproduce its observations between 17:00 and 18:30 UT. The second enhancement between 18:30 and 20:00 UT is not reproduced by any model. From about 20:00 onward BATS-R-US, GUMICS, and LFM again agree reasonably well with Wind. B_z measured by Wind has features similar

to B_z of Cluster 1 but with a magnitude of about one fourth of that of Cluster. For the whole event, BATS-R-US, GUMICS, and LFM agree with Wind quite well with one exception of about 30 min around 20:00 UT. OpenGGCM agrees well with Wind between 17:00 and 19:30 UT after which it starts to show very large variations. Again the B_z result of BATS-R-US, GUMICS, and LFM are similar to their respective results for B_z at Cluster 1 but with a 30 to 45 min delay.

[16] Figure 6 shows the cross polar cap potential (CPCP) in the northern hemisphere from simulations and calculations from SuperDARN along with the number of flow vectors that were used in the calculation as a function of time. The procedure of calculating SuperDARN CPCP [Ruohoniemi and Baker, 1998] consist of finding the best fit for the electric potential Φ from observed flow vectors using $\vec{v} = -\frac{(\nabla\Phi)\times\vec{B}}{B^2}$. In regions without data coverage, a statistical model based on the solar wind IMF [Ruohoniemi and Greenwald, 1996] is used to constrain the solution. Most of the time, over 100 flow vectors are available and, for example, starting at 21:30 UT, about 300 vectors are available for almost 2 h. When the number of available vectors is above about 100, they seem to be available on both the dawn and dusk side of the northern hemisphere (data not shown).

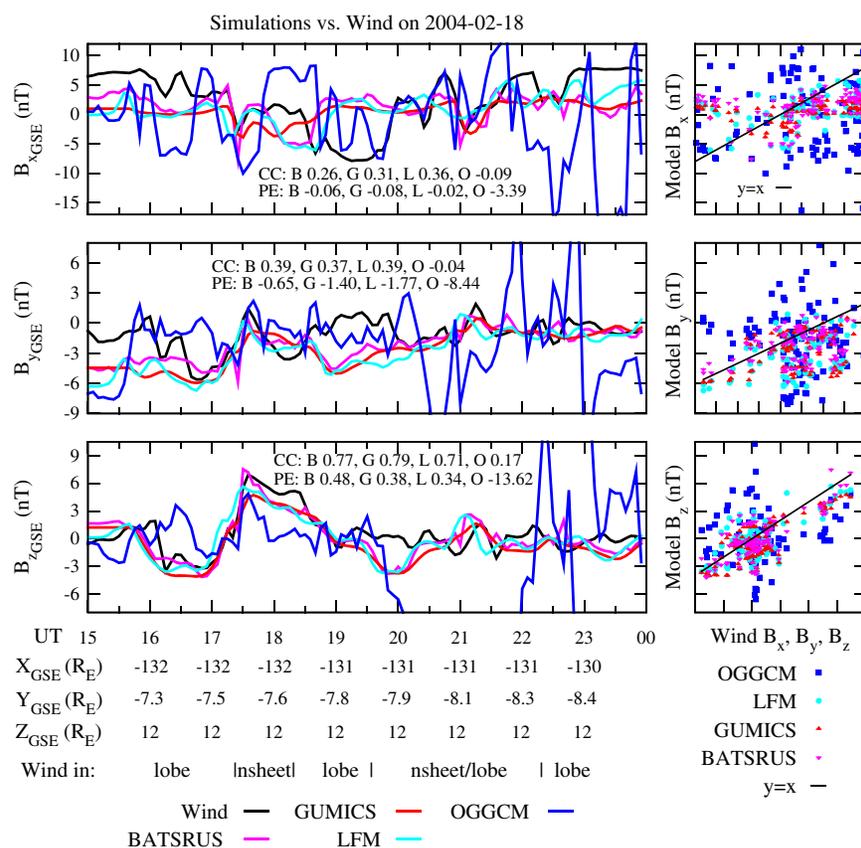


Figure 5. Magnetic field data from Wind and simulations shown in the same format as in Figure 3.

[17] During the event SuperDARN CPCP increases significantly four times with peak values at about 16:30, 19:30, 21:30, and 23:00 UT, which coincide with southward/northward turning of the IMF. A first-order minimum estimate for CPCP determined from the Defense Meteorological Satellite Program (DMSP) spacecraft available at NADIAWEB (http://cindispace.utdallas.edu/DMSP/NADIA_FAQ.html) gives values in the same range (~ 40 to ~ 100 kV) as SuperDARN. BATS-R-US follows observations most closely with significantly different values for a period of about 30 min only at 17:30 and 20:45 UT. GUMICS and LFM capture the dynamics of CPCP well, but their results differ from observations by a constant factor of about 0.7 and 3, respectively. OpenGGCM also captures the main behavior of CPCP but with smaller enhancements before 21:00 UT and a very large increase in CPCP at around 23:00 UT.

[18] Figure 7 shows the minimum distance of the magnetopause from the Earth within 30° from the Sun-Earth line (referred to as R_0 from hereinafter) from simulations and the empirical model of *Lin et al.* [2010] as a function of time. With the exception of the first 1.5 h, BATS-R-US and GUMICS show very similar results for the entire event, staying between 11 and 13 R_E . Their result is also quite close to the empirical model, but with an almost

constant offset of about 1.5 R_E . The results from LFM are closest to the empirical model with very good agreement (differences of less than 0.25 R_E) from about 17:00 to 20:30 UT and at other times the difference is at most about 1 R_E . In both, LFM and the empirical model, R_0 stays almost completely between 9 and 11 R_E . OpenGGCM has the lowest average value of R_0 of 8 R_E , and its dynamic range is much larger than the other models, varying between 6 and 12 R_E .

[19] The topology of the magnetic field in a global MHD simulation can give significant insight into the solution that was obtained and is essential for example when studying reconnection in a global setting [see e.g., *Dorelli et al.*, 2007]). Figures 8 and 9 show the magnetic field topology [*Rastätter et al.*, 2012] from simulations in the $y = 0 R_E$ plane at about 21:20 UT (tracing parameters in Figure 8: $N_1 = N_2 = 11$, adaptation = 6; flow line start positions in Figure 9: uniform random in cut plane). In Figure 8, traced magnetic field lines connected to the Earth at both ends are shown in red, field lines connected only to the northern or southern hemisphere are shown in yellow or green, respectively, and field lines not connected to the Earth are shown in blue, while in Figure 9, field lines connected to either one hemisphere are shown in black. In GUMICS two large plasmoids form in the magnetotail starting at

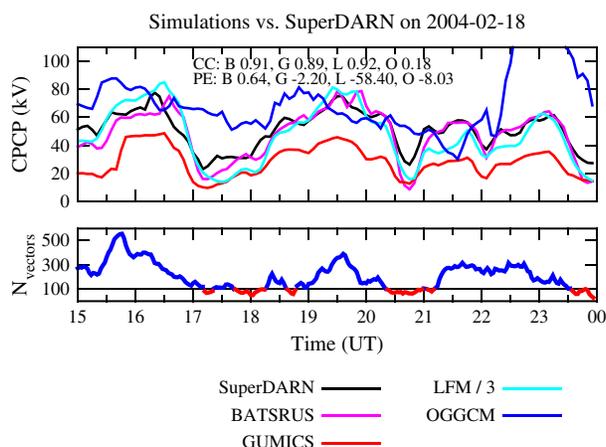


Figure 6. Five minute cross polar cap potential data from SuperDARN, obtained by averaging over a 5 min window, and CPCP data from four global MHD simulations as a function of time. Note that here LFM CPCP has been divided by 3. Correlation coefficients and prediction efficiencies between model results and observations are also shown, see section 5 for details. Also shown are the number of flow vectors that were used for calculating SuperDARN CPCP as a function of time.

20:30 and 23:30 UT, which is most likely caused by multiple large and fast rotations of the interplanetary magnetic field (IMF) clock angle [Honkonen *et al.*, 2011]. Figures 8b and 9b show the magnetic field topology from GUMICS at 21:12 UT, about 5 min before the first plasmoid dissipates. The plasmoid is visible as a region of closed magnetic field located in the far tail beyond $-100 R_E$. In BATS-R-US, the plasmoid forms about 10 min later than in GUMICS and is shown in Figures 8a and 9a 5 min before it dissipates. The result looks similar to that of GUMICS with a complicated structure of closed, lobe, and solar wind field lines in the ecliptic plane surrounded by lobe field lines above and below. Figures 8c and 9c show the plasmoid in LFM 5 min before it dissipates. The plasmoid forms at about the same time as that in BATS-R-US, but its structure is different. The closed field line regions stay closer to the ecliptic plane and do not detach from the Earth as in the case of BATS-R-US and GUMICS. At the time of plasmoid formation in other models, OpenGGCM does not show a closed field line region extending downstream from the Earth, but northern lobe field lines do show an additional region further down the tail. OpenGGCM shows significant changes in magnetic field topology prior to 20:00 UT. In BATS-R-US and LFM, the boundary between lobe and solar wind field lines in the magnetosheath is quite wavy, which is also noticeable in GUMICS.

5. Analysis

[20] In order to get a quantitative estimate for the performance of different models, the correlation coefficients

(CC) and prediction efficiencies (PE) with respect to observations were calculated. Table 2 presents the correlation coefficients between model predictions and measurements. For every model, the magnetic field component with the largest correlation coefficient for every spacecraft is shown in bold and the smallest coefficient in italics. In this section we use the expressions best correlated and best predicted only for describing values of the respective metrics of one magnetic field component relative to another of the same combination of spacecraft and model. As will be shown, in some cases, an average prediction is better than the modeled result for all components of B , but even then, one component will have the highest score to which we will refer to as best.

[21] Overall B_z for every spacecraft is most often (7/12) best correlated with measurements, but there are spatial differences. On the dayside (Cluster 1), the largest correlations are evenly distributed between B_x and B_z , while B_y has the lowest correlation for three models. On the nightside close to the Earth (Geotail), B_x is most often (3/4) best correlated with measurements while B_y has most often (3/4) the worst correlation. Far in the magnetotail (Wind), B_z is best correlated with measurements and B_x the worst for all models. B_y is never the best correlated magnetic field component for any model and spacecraft combination.

[22] When examining the correlations of each magnetic field component separately, several things can be observed. For three of the four models the highest correlation in B_x is obtained on the dayside and the lowest correlation far in the tail. In BATS-R-US and LFM the correlation of B_x decreases steadily from dayside to far tail. In GUMICS the correlation of B_x drops significantly from dayside to nightside and increases slightly from nightside to the far tail, which is probably due to GUMICS having a smaller resolution at Geotail than the other models. In OpenGGCM the highest B_x correlation is on the

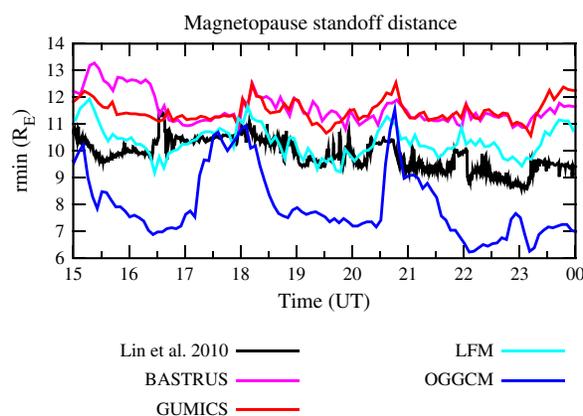


Figure 7. Minimum distance of the magnetopause from Earth within 30° from the Sun-Earth line in four global MHD simulations and the model of Lin *et al.* [2010] as a function of time.

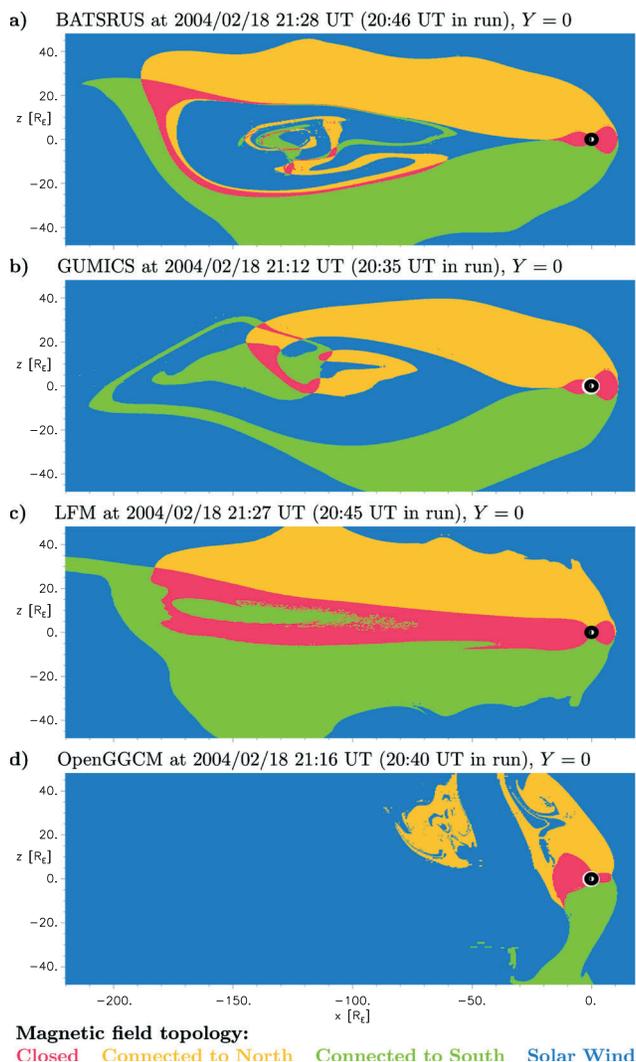


Figure 8. The color coded magnetic field topology from simulations in the $y = 0 R_E$ plane. Traced magnetic field lines connected to the Earth at both ends are shown in red, field lines connected only to the northern (southern) hemisphere are shown in yellow (green), field lines not connected to the Earth are shown in blue. (a) BATS-R-US at 21:28 UT, (b) GUMICS at 21:12 UT, (c) LFM at 21:27 UT, and (d) OpenGGCM at 21:16 UT.

nightside. For all models, the lowest correlations of both B_y and B_z are on the nightside close to Earth with only one exception (OpenGGCM B_y at Wind). Overall, 37 out of 40 correlation coefficients are positive when SuperDARN is included.

[23] The prediction efficiency (PE) for a discrete signal is calculated following *Pulkkinen et al.* [2011]:

$$PE = 1 - \frac{\langle (x_{obs} - x_{sim})^2 \rangle_i}{\sigma_{obs}^2}$$

where x_{obs} and x_{sim} are the observed and simulated signals, respectively, $\langle \dots \rangle_i$ indicates an arithmetic mean taken over i (i.e., time) and σ_{obs}^2 is the variance of the observed signal. A PE value of 1 indicates a perfect prediction while a PE value of 0 is equal to using the mean value of the signal as a predictor. The prediction efficiencies of the models are presented in Table 3. For every model the magnetic field component with the largest prediction efficiency for every spacecraft is shown in bold and the smallest prediction efficiency in italics.

[24] Overall, B_x for every spacecraft is most often (7/12) best predicted by the models but again there are spatial differences. On the dayside and nightside close to the Earth, B_x is predicted best by three of the four models, but in the far tail, the highest prediction efficiency is mostly (3/4) obtained for B_z instead. On the dayside and far in the tail, B_y is predicted the worst almost without exception, but on the nightside close to Earth, B_z is predicted worst by all models. All the models are worse than using an average value for predicting B_y with the exception of one model for only one spacecraft (GUMICS on the dayside). On the other hand, for three models, the B_x and B_z prediction efficiencies on the dayside are above 0.5 and above 0.3 for B_z in the far tail.

[25] When examining the prediction efficiencies of each magnetic field component separately, several things are observed: For BATS-R-US and LFM, the highest PE for B_x is on the dayside and the lowest in the far tail, a situation identical to the correlation coefficients of B_x for these models. For GUMICS-4 and OpenGGCM, the highest PE for B_x is also on the dayside, but the lowest one is on the nightside close to Earth. The PE of B_y decreases downstream for GUMICS-4 and OpenGGCM but for BATS-R-US and LFM, B_y PE has the highest value on the nightside close to Earth. For all the models, B_z PE is significantly lower on the nightside than either the dayside or the far tail. Only BATS-R-US predicts the CPCP better than an average value would, and all other models are worse than using a random value. Overall, 13 out of 40 prediction efficiencies are positive when SuperDARN is included.

[26] For the combined magnetic field CC and PE results from all models B_x has the largest value among all components 12/24 times while B_y has the largest value 1/24 times and B_z 22/24 times. The number of times B_x , B_y and B_z have the smallest value among components are 5/24, 13/24 and 6/24 respectively.

[27] In order to estimate the effect that different modules/parameters of a model can have on the simulation result, three runs were done with a newer version of the BATS-R-US model (version 20110131) with different parameters: (1) only the version was changed, (2) higher resolution was used in the magnetosphere, and (3) the Rice Convection Model (RCM) module was included which solves the adiabatic drift of isotropic particle distributions [*Toffoletto et al.*, 2005] in the inner magnetosphere and provides density and pressure corrections to the magnetospheric module [*Tóth et al.*, 2005]. At Cluster 1, the standard resolution run has a resolution of $0.5 R_E$ (up

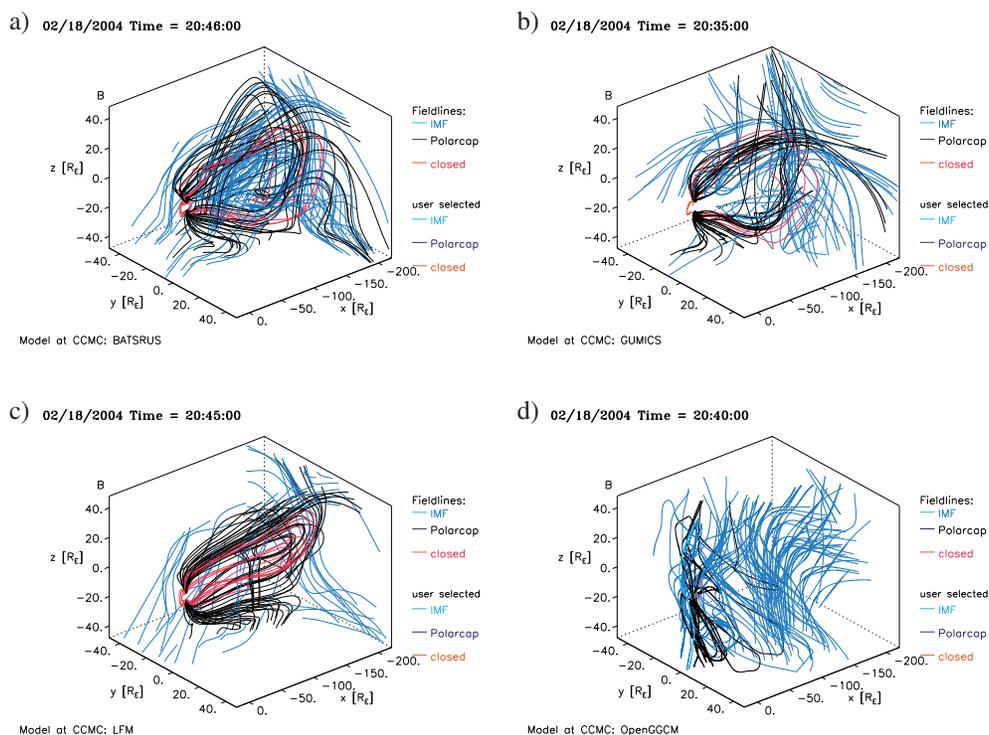


Figure 9. Magnetic field lines traced from the $y = 0 R_E$ plane in simulations shown at same times as in Figure 8. Field lines connected to the Earth at both ends are shown in red, field lines connected only to the northern or southern hemisphere are shown in black, field lines not connected to the Earth are shown in blue.

to about $8 R_E$ distance from Earth) while the high resolution version has a resolution of $0.25 R_E$ (up to about $16 R_E$ distance from Earth on the dayside). At Geotail the resolutions are $1 R_E$ and $0.5 R_E$, respectively although in the high resolution run the resolution decreases to $1 R_E$ when $Z > 6$ and further to $2 R_E$ when $Z > 12 R_E$.

[28] We only summarize the results for CC and PE in these runs, but the simulation results are again available through CCMC (run numbers 112112_1, 112112_2, 112112_3, and 112312_1 all prefixed with

“Ilja_Honkonen_”). The results for B_x at Cluster orbit do not differ significantly between any version or module combination of BATS-R-US that was tested. At Geotail only RCM improved the results for B_x noticeably by increasing the absolute value of both CC and PE by about 0.1. For B_y , the differences are less straightforward: CC at Cluster improved slightly (about 0.05) with the new version regardless of the requested resolution and further by about 0.05 when using RCM. At Geotail, the new version improved CC of B_y by 0.15 which higher resolution

Table 2. Correlation Coefficients Between Model Magnetic Fields, Cross Polar Cap Potential, and Measurements for the Event of 18 Feb 2004^a

		BATS-R-US	GUMICS-4	LFM	OpenGGCM
Cluster 1	B_x	0.89	0.83	0.86	<i>0.05</i>
	B_y	0.63	<i>0.62</i>	<i>0.58</i>	0.25
	B_z	0.88	0.92	0.85	0.53
Geotail	B_x	0.70	0.24	0.63	0.20
	B_y	<i>0.24</i>	<i>-0.24</i>	<i>0.20</i>	0.09
	B_z	0.58	0.59	0.56	<i>0.06</i>
Wind	B_x	0.26	0.31	0.36	<i>-0.09</i>
	B_y	0.39	0.37	0.39	<i>-0.04</i>
	B_z	0.77	0.79	0.71	0.17
SuperDARN	CPCP	0.91	0.89	0.92	0.18

^aFor every model, separately, the magnetic field component with the largest correlation coefficient for every spacecraft is shown in bold and the smallest coefficient in italics.

Table 3. Magnetic Field and Cross Polar Cap Potential Prediction Efficiencies of Global MHD Models for the Event of 18 Feb 2004^a

		BATS-R-US	GUMICS-4	LFM	OpenGGCM
Cluster 1	B_x	0.78	0.53	0.69	-0.23
	B_y	<i>-0.16</i>	<i>0.26</i>	<i>-1.14</i>	<i>-4.78</i>
	B_z	0.69	0.85	0.61	-1.81
Geotail	B_x	0.45	-1.67	0.25	-3.81
	B_y	-0.07	-0.84	-0.35	-8.23
	B_z	-3.74	-3.13	<i>-1.50</i>	<i>-54.95</i>
Wind	B_x	-0.06	-0.08	-0.02	-3.39
	B_y	<i>-0.65</i>	<i>-1.40</i>	<i>-1.77</i>	<i>-8.44</i>
	B_z	0.48	0.38	0.34	-13.62
SuperDARN	CPCP	0.64	-2.20	-58.40	-8.03

^aFor every model, separately, the magnetic field component with the largest prediction efficiency for every spacecraft is shown in bold and the smallest efficiency in italics.

increased further by 0.06. Interestingly in this case, including RCM decreased CC by about 0.05. The prediction efficiency of B_y at Cluster improved significantly for the new version (by 0.35) and increasing the resolution and including RCM further improved the result by 0.1 and 0.2, respectively. At Geotail, the new version increased PE by almost 0.2 while increasing the resolution or including RCM decreased the result slightly. For B_z the results are relatively straightforward: A new version of BATS-R-US does not affect either CC or PE significantly; increasing the resolution improves all results noticeably at Cluster while having no or insignificant effect at Geotail. Interestingly, using RCM gives the worst results for B_z everywhere except for PE at Geotail, but there the results for all tested models are already worse than a random prediction.

6. Discussion

6.1. Metrics

[29] In this work the prediction efficiency (PE) and correlation coefficient (CC) metrics are used to obtain a quantitative estimate on model performance. An intuitive picture of these metrics can be obtained from the cross polar cap potential (CPCP) results where CC seems to quantify how well a model reproduces the “dynamics” of observations while PE indicates the quality of predicting the absolute values of observations. The CPCP prediction of LFM has the highest CC score of all models and indeed the relative changes of LFM CPCP correspond best to SuperDARN observations. On the other hand, the PE score of LFM CPCP is by far the lowest, which is not surprising given that LFM CPCP is mostly a factor of 3 larger than that of SuperDARN or any other model. As stated by *Pulkkinen et al.* [2011], no one metric is the absolute best and the choice of the metric depends on the situation. For some applications it could be a valid, albeit an unphysical, approach to divide the LFM CPCP by three in order to obtain the best available prediction for CPCP based on the CC metric.

[30] Overall, based on the PE metric, none of the tested models seem good at predicting observations on the nightside at a distance of about 25 R_E or more from the Earth. With the exception of B_x at Geotail and B_z at Wind, all model predictions at those satellites are worse than using an average value for the prediction, and even in the rest of the cases, PE is less than 0.5. On the dayside closer to Earth at about 14 R_E , all model PEs are significantly higher with the PE of B_x and B_z being over 0.5 for three out of four models. When the CC metric is used, all models fare substantially better, and interestingly, the highest and lowest values of CC occur in about the same locations and for the same components of B as with PE.

[31] In BATS-R-US and LFM the values of CC and PE tend to decrease steadily from the dayside to the far tail with the exception of B_z . For a system with high Mach number(s) flow past an obstacle, it is reasonable that model predictions are most accurate upstream of the obstacle and decrease downstream from there since turbulence and other effects have had more time to affect the system. In GUMICS and OpenGGCM, both CC and PE in the nightside close to the Earth often have smaller values than in the far tail. There does not seem to be a simple explanation for this behavior since GUMICS as a model is closer to BATS-R-US than LFM is to BATS-R-US (Table 1), but it is GUMICS that behaves differently from BATS-R-US in this respect.

[32] B_z is the largest exception to the above rule that metric scores decrease steadily downstream from the dayside since in all models and for both CC and PE, the score is higher in the far tail than in the nightside close to Earth. The lack of modeled physics in the nightside does not seem to explain this since including RCM in BATS-R-US decreases the nightside CC and PE scores if they are positive to begin with. One obvious explanation for this behavior could be the fact the axis of Earth’s strong intrinsic dipole field is also directed in the general direction of the GSE Z axis. In order to verify this, a similar comparison would probably have to be carried out for Neptune’s magnetosphere where the dipole axis can point almost directly sunward [*Ness et al.*, 1989].

6.2. Scatter Plots

[33] The scatter plots in Figures 3 to 5 illustrate the quality of predictions from another point of view. They allow one to understand the metric scores calculated here better and also let us estimate a limits to the CC and PE scores above which the simulation results can be considered good. For example, it is more apparent from the scatter plot than the time series that all models underestimate B_z at Geotail quite strongly. In a BATS-R-US run with RCM included, the shape of the point distribution does not change significantly but the whole distribution moves closer to the diagonal. Based on visual inspection of scatter plots, the modeled results seem good when both CC and PE are above about 0.6. Also using only the CC metric for estimating quality of the result can be misleading as the points in a scatter plot can still be quite far from the diagonal (e.g., GUMICS B_x at Cluster or BAST-R-US/LFM B_x at Geotail).

6.3. Cross Polar Cap Potential

[34] The CPCP result of GUMICS differs from that of SuperDARN by a constant factor of about 0.7 which can be explained, at least partially, by magnetospheric resolution and the dipole tilt angle. When using the Run-on-Request system, if the tilt angle is not updated with time, its direction is set to the start time of the simulation. In a previous simulation of this event with GUMICS, the dipole tilt angle was set to its average value during the simulated event [Honkonen *et al.*, 2011] and the CPCP prediction was noticeably higher, i.e., closer to SuperDARN and the other models. Also, increasing the resolution of the inner magnetosphere in GUMICS gives higher field-aligned currents (FAC), which increase CPCP further.

[35] The CPCP from SuperDARN reaches its saturation point of about 80 kV [e.g., Shepherd *et al.*, 2003] twice during the event at 16:30 and 19:30 UT. In this case the possible saturation of CPCP most likely would not change the relative result between models significantly because at those times all model predictions are less than or equal to SuperDARN with the exception of LFM, which is a factor of three higher. At times the number of flow vectors used for calculating SuperDARN CPCP falls below 100, and the number does not increase much beyond 300. This may cast some doubt on the calculated CPCP since, for example, the large statistical study of Grocott *et al.* [2012] only included periods with 300 or more flow vectors. We argue that the number of vectors and hence the reliability of SuperDARN CPCP is adequate for the purpose of comparing global MHD models, with quite different CPCP predictions, to observations.

6.4. Additional Physics

[36] Pulkkinen *et al.* [2011] reported that neither increasing spatial resolution in OpenGGCM nor including thermospheric physics in LFM systematically improved the performance of either model with respect to ground magnetic field observations. In this event increasing the resolution in BATS-R-US has a significant effect only on

the dayside CC and PE of B_z , and in particular, higher resolution does not have a significant effect on B_x anywhere. Contrary to Pulkkinen *et al.* [2011], including the RCM module in BATS-R-US does not lead to an improved result in this case. Although the result for B_x does improve on the nightside by including RCM, the result for B_z becomes worse on the dayside. The reason for this behavior would be difficult to pinpoint based on even several tests. The possibilities range from small mistakes in the code, installation or usage to fundamental problems in the representation of the physics in each separate model, or in their coupling together. It is clear that including additional physical models in a simulation does not automatically guarantee a better result in the whole simulated volume.

6.5. Magnetopause Standoff Distance

[37] The magnetopause standoff distance between different MHD models varies by almost $6 R_E$ at 16:30 and 22:00 UT while at 18:00 UT all models show an almost identical value and are quite close at 20:30 UT. The standoff distance from an empirical model tends to fall in the middle of MHD models except from about 21:00 UT onward where the standoff distance is lower than in all but one MHD model. While proper validation would require a comparison to observations that is not possible in this case using the current CCMC interface. Nevertheless, a comparison to an empirical model shows what values and dynamical behavior to expect based on the upstream solar wind conditions.

[38] The increase in standoff distance of all models at 18:00 UT is probably due to the sudden large decrease in solar wind density while the increase at 20:30 UT might be due to northward turning of IMF B_z . The empirical model shows similar behavior although the increase at 18:00 UT is smaller than in MHD. At 22:00 UT the standoff distance again increases, probably due to northward turning of IMF B_z , in BATS-R-US, GUMICS, LFM, and the empirical model. While the standoff distance increases several times to a very similar value between all MHD models, the standoff distance seems to subsequently return to a baseline value that is different for each model. Finding the cause of this will require further investigation in subsequent works, as there is no apparent explanation for this behavior in the upstream solar wind conditions.

6.6. Statistical Studies Needed

[39] When validating, verifying, or just comparing models, as many parameters as possible should be kept constant. Unfortunately, this is difficult to accomplish with the models used here especially through the CCMC interface, which limits the parameter space of models available to users. For example, even when a higher resolution run of BATS-R-US is requested, the resolution does not increase in the whole simulation domain but is lower, for example, around the lobes. As shown in Table 1, there are also significant differences between, e.g., the magnetospheric grid used by different models. In this work we examined only event, and more may be needed in

order to draw solid conclusions on the performance of global MHD models in the Earth's magnetosphere. Due to the complicated physics involved and the differences in global models, it would be important to not only simulate single events but to run weeks or even many months worth of simulations in order to assess model performance using various metrics as a function of, for example, AE (<http://wdc.kugi.kyoto-u.ac.jp/aedir>) and Dst indices, the upstream solar wind driver, substorm phase, etc. This would allow the users of global models to estimate the quality of the solution for a particular event and could provide weights for statistical comparisons of simulations and observations. For model developers it would provide, for example, a baseline quality against which various modifications and changes in model parameters could be compared using a smaller but representative set of events.

[40] As a final note, there are four different global MHD models available at CCMC through a consistent interface, and each user can start several runs daily. Based on the results presented here and in previous works, there is virtually no reason to limit oneself to only one model when simulating the Earth's magnetosphere using CCMC resources. If two or more independent models agree on a particular result, it is very likely as close to reality as current state-of-the-art in global MHD can reasonably get.

6.7. Large Plasmoid Formation in Global MHD

[41] The results presented in Figures 8 and 9 lend more credibility to the hypothesis put forward by *Honkoken et al.* [2011] that multiple large and fast rotations of the IMF clock angle result in large plasmoid formation in a global MHD simulation. In three out of four different global MHD models, two large plasmoids form in the magnetotail during the event, and the plasmoids occur close in time between all three models in both cases although there are 5 to 15 min differences between the three models in the stages of plasmoid formation. The plasmoid structure is most similar between BATS-R-US and GUMICS with LFM also showing the closed magnetic field line region extending about $-200 R_E$ downstream from Earth.

[42] The three models that agree on plasmoid formation also exhibit a large cross polar cap potential drop prior to the downstream growth of the closed magnetic field line region. GUMICS shows only very small variations in the ionospheric conductivities during the event, and hence changes to CPCP are almost completely due to changes in the FACs. Although the FACs in BATS-R-US vary by more than a factor of 2, the conductivities in the auroral oval also change moderately when the CPCP varies. In LFM both FACs and conductivities change significantly while in OpenGGCM only conductivities change drastically during variations of CPCP. Thus it seems that in order for a large tail plasmoid to be formed in a global MHD simulation, significant changes in ionospheric FACs are required and that ionospheric conductivities can have a significant effect on the structure of the formed plasmoid.

7. Conclusions

[43] In this work the performance of four global MHD models (BATS-R-US, GUMICS-4, LFM, and OpenGGCM) in the Earth's magnetosphere is studied by comparing model predictions to the magnetic field measurements of Cluster 1, Geotail, and Wind spacecraft during a multiple substorm event. Model results for the cross polar cap potential are also compared to the measurements of SuperDARN, and the model magnetopause standoff distances are compared to the empirical magnetopause model of *Lin et al.* [2010]. All simulations are executed through the CCMC Run-on-Request system. Comparisons are conducted using two quantitative and objective metrics: correlation coefficient and prediction efficiency. We find that for all four models, the best performance is on the dayside and, generally, model performance decreases steadily downstream from the Earth. From different components of the magnetic field, B_x is most often best predicted and correlated both on the dayside and the nightside close to the Earth whereas in all models B_z CC and PE are substantially higher in the far tail than for other components. On the dayside CCs are above 0.5 most of the time with B_x and B_z CCs close to 0.9 for three out of four models. In the magnetotail at a distance of about $130 R_E$, the prediction efficiency of all models is below that of using an average value for the prediction with the exception of B_z . We also find that increasing the resolution or coupling an additional physical model does not automatically increase model performance at least with respect to the CC and PE metrics. With a coupled inner magnetosphere module, the performance of BATS-R-US increases significantly close to the Earth for B_y and in all relevant cases decreases moderately for B_z .

[44] **Acknowledgments.** This work is a part of the project 200141-QuESpace, funded by the European Research Council under the European Community's seventh framework programme. The work of I.H. and M.P. is supported by project 218165 of the Academy of Finland. A.G. is supported by NERC Grant NE/G019665/1. The National Center for Atmospheric Research is supported by the National Science Foundation. We thank the rest of the CCMC staff for providing a valuable service, the Cluster Active Archive, Coordinated Data Analysis Web and the instrument teams and PIs of Geotail MGF, Wind MFI, Cluster FGM, and ACE MAG and SWEPAM (S. Kokubun, R. Lepping, A. Balogh, N.F. Ness, D.J. McComas, respectively) for providing the spacecraft data used in this study. We also thank the World Data Center for Geomagnetism for the Kyoto AE and Dst index services and the anonymous referee for insightful comments. I.H. thanks C. Anekallu for insightful discussions.

References

- Aikio, A. T., T. Pitkänen, I. Honkoken, M. Palmroth, and O. Amm (2013), IMF effect on the polar cap contraction and expansion during a period of substorms, *Ann. Geophys.*, submitted.
- Balogh, A., et al. (1997), The cluster magnetic field investigation, *Space Sci. Rev.*, 79, 65–91, doi:10.1023/A:1004970907748.
- Brackbill, J. U., and D. C. Barnes (1980), The effect of nonzero product of magnetic gradient and B on the numerical solution of the magnetohydrodynamic equations, *J. Comput. Phys.*, 35, 426–430.
- Chisham, G., et al. (2007), A decade of the Super Dual Auroral Radar Network (SuperDARN): Scientific achievements, new techniques

- and future directions, *Surv. Geophys.*, 28(1), 33–109, doi:10.1007/s10712-007-9017-8.
- Dorelli, J. C., A. Bhattacharjee, and J. Raeder (2007), Separator reconnection at Earth's dayside magnetopause under generic northward interplanetary magnetic field conditions, *J. Geophys. Res.*, 112, A02202, doi:10.1029/2006JA011877.
- Grocott, A., S. E. Milan, S. M. Imber, M. Lester, and T. K. Yeoman (2012), A quantitative deconstruction of the morphology of high-latitude ionospheric convection, *J. Geophys. Res.*, 117, A05317, doi:10.1029/2012JA017580.
- Honkoniemi, I., S. von Althaus, A. Sandroos, P. Janhunen, and M. Palmroth (2013), Parallel grid library for rapid and flexible simulation development, *Comput. Phys. Commun.*, 184, 1297–1309, doi:10.1016/j.cpc.2012.12.017.
- Honkoniemi, I., M. Palmroth, T. I. Pulkkinen, P. Janhunen, and A. Aikio (2011), On large plasmoid formation in a global magnetohydrodynamic simulation, *Ann. Geophys.*, 29(1), 167–179, doi:10.5194/angeo-29-167-2011.
- Janhunen, P., M. Palmroth, T. Laitinen, I. Honkoniemi, L. Juusola, G. Facskó, and T. I. Pulkkinen (2012), The GUMICS-4 global MHD magnetosphere-ionosphere coupling simulation, *J. Atmos. Sol. Terr. Phys.*, 80, 48–59, doi:10.1016/j.jastp.2012.03.006.
- Kokubun, S., T. Yamamoto, M. H. Acuña, K. Hayashi, K. Shiokawa, and H. Kawano (1994), The GEOTAIL magnetic field experiment, *J. Geomag. Geoelectr.*, 46, 7–21.
- Krausmann, E., and V. Bothmer (2012), Meeting report: European Commission's space-weather awareness dialogue, *Space Weather*, 10, S04006, doi:10.1029/2012SW000790.
- Ledvina, S. A., Y.-J. Ma, and E. Kallio (2008), Modeling and simulating flowing plasmas and related phenomena, *Space Sci. Rev.*, 139(1-4), 143–189, doi:10.1007/s11214-008-9384-6.
- Lepping, R. P., et al. (1995), The WIND magnetic field investigation, *Space Sci. Rev.*, 71, 207–229, doi:10.1007/BF00751330.
- Lin, R. L., X. X. Zhang, S. Q. Liu, Y. L. Wang, and J. C. Gong (2010), A three-dimensional asymmetric magnetopause model, *J. Geophys. Res.*, 115, A04207, doi:10.1029/2009JA014235.
- 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(15-16), 1333–1350, doi:10.1016/j.jastp.2004.03.020.
- McPherron, R. L. (1991), *Physical Processes Producing Magnetospheric Substorms and Magnetic Storms*, Geomagnetism, vol. 4, J. Jacobs, Academic Press Ltd., London, England.
- Merkin, V. G., and J. G. Lyon (2010), Effects of the low-latitude ionospheric boundary condition on the global magnetosphere, *J. Geophys. Res.*, 115, A10202, doi:10.1029/2010JA015461.
- Ness, N. F., M. H. Acuña, L. F. Burlaga, J. E. P. Connerney, R. P. Lepping, and F. M. Neubauer (1989), Magnetic fields at Neptune, *Science*, 246, 1473–1478, doi:10.1126/science.246.4936.1473.
- 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, doi:10.1006/jcph.1999.6299.
- 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., D. Larson, W. Li, E. L. Kepko, and T. Fuller-Rowell (2008), OpenGGCM simulations for the THEMIS mission, *Space Sci. Rev.*, 141(1-4), 535–555, doi:10.1007/s11214-008-9421-5.
- Rastätter, L., et al. (2013), Geospace environment modeling 2008–2009 challenge: Dst index, *Space Weather*, accepted, 11(4), 187–205, doi:10.1002/swe.20036.
- Rastätter, L., M. M. Kuznetsova, D. G. Sibeck, and D. H. Berrios (2012), Scientific visualization to study flux transfer events at the Community Coordinated Modeling Center, *Adv. Space Res.*, 49(11), 1623–1632, doi:10.1016/j.asr.2011.12.034.
- 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.
- Ruohoniemi, J. M., and K. B. Baker (1998), Large-scale imaging of high-latitude convection with Super Dual Auroral Radar Network HF radar observations, *J. Geophys. Res.*, 103, 20797–20811, doi:10.1029/98JA01288.
- Ruohoniemi, J. M., and R. A. Greenwald (1996), Statistical patterns of high-latitude convection obtained from Goose Bay HF radar observations, *J. Geophys. Res.*, 101, 21743–21763, doi:10.1029/96JA01584.
- Ryu, D., and T. W. Jones (1995), Numerical magnetohydrodynamics in astrophysics: Algorithm and tests for one-dimensional flow, *The Astrophys. J.*, 442, 228–258, doi:10.1086/175437.
- Shepherd, S. G., J. M. Ruohoniemi, and R. A. Greenwald (2003), Testing the Hill model of transpolar potential with Super Dual Auroral Radar Network observations, *Geophys. Res. Lett.*, 30(1), 1002, doi:10.1029/2002GL015426.
- Stone, E. C., A. M. Frandsen, R. A. Mewaldt, E. R. Christian, D. Margolies, J. F. Ormes, and F. Snow (1998), The advanced composition explorer, *Space Sci. Rev.*, 86, 1–22, doi:10.1023/A:1005082526237.
- Tanaka, T. (1994), Finite volume TVD scheme on an unstructured grid system for three-dimensional MHD simulation of inhomogeneous systems including strong background potential fields, *J. Comput. Phys.*, 111, 381–389.
- Toffoletto, F., S. Sazykin, R. Spiro, and R. Wolf (2005), Inner magnetospheric modeling with the Rice Convection Model, *Space Sci. Rev.*, 107(1-2), 175–196, doi:10.1023/A:1025532008047.
- 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.
- Viljanen, A., E. I. Tanskanen, and A. Pulkkinen (2006), Relation between substorm characteristics and rapid temporal variations of the ground magnetic field, *Ann. Geophys.*, 24(2), 725–733, doi:10.5194/angeo-24-725-2006.
- Welling, D. T., J. Koller, and E. Camporeale (2012), Verification of Spacepy's radial diffusion radiation belt model, *Geosci. Model Dev.*, 5(2), 277–287, doi:10.5194/gmd-5-277-2012.