SPACE WEATHER, VOL. 5, S01005, doi:10.1029/2006SW000222, 2007

Predicting magnetopause crossings at geosynchronous orbit during the Halloween storms

R. E. Lopez,[1] S. Hernandez,[1] M. Wiltberger,[2] C.-L. Huang,[3] E. L. Kepko,[3] H. Spence,[3] C. C. Goodrich,[3] and J. G. Lyon[4]

Received 18 January 2006; revised 1 September 2006; accepted 5 September 2006; published 16 January 2007.

[1] In late October and early November of 2003, the Sun unleashed a powerful series of events known as the Halloween storms. The coronal mass ejections launched by the Sun produced several severe compressions of the magnetosphere that moved the magnetopause inside of geosynchronous orbit. Such events are of interest to satellite operators, and the ability to predict magnetopause crossings along a given orbit is an important space weather capability. In this paper we compare geosynchronous observations of magnetopause crossings during the Halloween storms to crossings determined from the Lyon-Fedder-Mobarry global magnetohydrodynamic simulation of the magnetosphere as well to predictions of several empirical models of the magnetopause position. We calculate basic statistical information about the predictions as well as several standard skill scores. We find that the current Lyon-Fedder-Mobarry simulation of the storm provides a slightly better prediction of the magnetopause position than the empirical models we examined for the extreme conditions present in this study. While this is not surprising, given that conditions during the Halloween storms were well outside the parameter space of the empirical models, it does point out the need for physics-based models that can predict the effects of the most extreme events that are of significant interest to users of space weather forecasts.

Citation: Lopez, R. E., S. Hernandez, M. Wiltberger, C.-L. Huang, E. L. Kepko, H. Spence, C. C. Goodrich, and J. G. Lyon (2007), Predicting magnetopause crossings at geosynchronous orbit during the Halloween storms, *Space Weather*, *5*, S01005, doi:10.1029/2006SW000222.

1. Introduction

[2] As reliance on spacecraft technology increases, our society becomes more vulnerable to space weather [*Carlowicz and Lopez*, 2002]. Thus an increasing amount of research is being done in the field of space weather [*Lanzerotti*, 2003] in order to be able to predict the space environment and the configuration of Earth's magnetosphere. In the past few years the Center for Integrated Space weather Modeling (CISM), has been funded to do fundamental research on coupled models that will extend from the Sun to the Earth [*Hughes and Hudson*, 2004]. Some of the models are empirical [e.g., *Siscoe et al.*, 2004], while others are based on a numerical simulation approach [e.g., *Elkington et al.*, 2004]. All of the models are to be tested against a set of metrics [*Spence et al.*, 2004], for both "operational" and "science" components, although some of the latter may also be of interest to operations that can be impacted by space weather.

[3] One metric that is of interest to satellite operators is the question of whether their spacecraft will exit the magnetosphere during an event. According to staff at the Space Environment Center, this issue is of interest to spacecraft operations (T. Onsager, personal communication, 2006). While there are excellent empirical models of the magnetopause position that could be used to determine if a satellite will exit the magnetosphere, the periods of special interest are during times when empirical models are outside their range of validity because of extreme solar wind conditions. Therefore it seems reasonable to use a physics-based simulation model to predict magnetopause crossings during such events, and to document the ability of the physics-based model to make accurate predictions.

[4] Previous work by *Shue et al.* [2000], *Yang et al.* [2002], and *Dmitriev et al.* [2004] compared several empirical

[1]Department of Physics and Space Sciences, Florida Institute of Technology, Melbourne, Florida, USA.
[2]High-Altitude Observatory, National Center for Atmospheric Research, Boulder, Colorado, USA.
[3]Department of Astronomy, Boston University, Boston, Massachusetts, USA.
[4]Department of Physics and Astronomy, Dartmouth College, Hanover, New Hampshire, USA.

magnetopause models to observations, and quantified these comparisons by calculating skills scores of the kind that are used in the meteorological community. In this paper we examine the accuracy of the predictions of such models during an extreme event along with predictions made by a numerical simulation model of the magnetosphere that is one of the four core models in the CISM portfolio. This will establish a baseline for the CISM numerical simulation model of the magnetosphere that may be used in future comparisons to determine if later versions of the numerical simulation are producing better predictions during extreme events.

2. Modeling the Magnetopause Position

[5] Over the years, a number of models of the magnetopause have been presented in the literature [e.g., *Roelof and Sibeck*, 1993; *Petrinec and Russell*, 1993; *Shue et al.*, 1997, 1998]. Generally such models are obtained by using observations of magnetopause crossings by spacecraft along with simultaneous solar wind data to fit analytic functions representing magnetopause position and determine the coefficients of those functions. The model of *Petrinec and Russell* [1993] is somewhat different in that it uses observations of the magnetic field in the lobe and MHD pressure balance to calculate the magnetopause shape. The range of solar wind parameters for which the models cited here are valid varies among the models. *Roelof and Sibeck's* [1993] model is valid for $B_z = \pm 5$ nT, *Petrinec and Russell's* [1993] model is valid for $B_z = \pm 10$ nT, and the model of *Shue et al.* [1998] is valid for $B_z = \pm 18$ nT. However, even the largest of these B_z ranges can be exceeded by a large magnetic cloud. Similarly, the dynamic pressure one can find in a magnetic cloud (especially the fast ones) can be well above the ranges of dynamic pressure for which the empirical models are valid.

[6] Extreme conditions outside the valid ranges of empirical models are of great interest to the space weather community since it is during extreme events that one would expect the most hazardous environmental conditions in space. On the other hand, physics-based numerical simulation models should, in principle, be as valid during extreme conditions as they are during quiet times. Therefore it is reasonable to compare a physics-based numerical simulation model to empirical models during an extreme event outside of the parameter range of the empirical models to see if the physics-based model is able to provide a better prediction than the empirical models. The Halloween storms provide just such an extreme event for testing the ability of a physics-based numerical simulation model to provide predictions of the magnetopause position during a period when one would not want to use an empirical model.

3. Halloween Storms and Their Simulation

[7] The early part of October 2003 seemed uneventful for the point of view of solar activity. All that changed on 18 October when sunspot group 484 appeared. Two other large sunspot groups soon joined 484. Over the next few days these magnetically intense regions were the sites of a series of energy releases, including the most powerful flare yet recorded by modern instruments [*Lopez et al.*, 2004]. The events were of considerable interest both to the scientific community and to the space weather community [*Onsager et al.*, 2004].

[8] To simulate the event, we use the Lyon-Fedder-Mobarry (LFM) code, which is a fully three-dimensional magnetohydrodynamic simulation of the solar wind–magnetosphere interaction [e.g., *Lyon et al.*, 2004]. The simulation uses solar wind data as an outer boundary condition (the inner boundary being a 2-D semiempirical ionosphere model), so it is able to model real events [e.g., *Lopez et al.*, 2000] or be used as an experimental tool to study solar wind–magnetosphere coupling [e.g., *Wiltberger et al.*, 2003]. LFM is one of the basic building blocks of the overall CISM space environment simulation [*Goodrich et al.*, 2004; *Wiltberger et al.*, 2004], and it has been used successfully in the past to model periods of strong driving of the magnetosphere by the solar wind as one finds during large magnetic storms [*Lopez et al.*, 2000].

[9] The simulation of the Halloween storms presents a challenge because the solar wind information is incomplete because of the fact that ACE suffered a loss of plasma density data. However, we reconstructed a solar wind data file by using densities inferred from the Geotail plasma wave experiment (Geotail was upstream of the Earth's bow shock throughout the event). Those data were shifted back in time to align with the ACE data (using the ACE plasma velocities). A merged data file was created and used to drive the LFM simulation of the event. *Dmitriev et al.* [2005] made an evaluation of a similarly reconstructed solar wind data file and found that the Geotail plasma wave data was a relatively good measure of the solar wind densities, except for 1600–1800 UT on 29 October, 1700–1800 UT on 30 October, and 0000–0400 UT on 31 October, when they inferred that the solar wind dynamic pressure was larger than one would calculated on the basis of the densities inferred from Geotail.

[10] The reconstructed solar wind data, propagated to 30 R_E upstream of the Earth (the earthward edge of the LFM simulation grid), are presented in Figure 1. Typically the X component is expressed as a linear function of Y and Z so that the full 3-D field can be propagated into the simulation while still preserving a zero divergence of the magnetic field. However, in this case the simulation the solar wind X component of the magnetic field was set to zero. The variation in the dipole tilt angle is included by running the simulation in SM coordinates, with the solar wind data rotated into this magnetically aligned coordinate system.

[11] Geosynchronous orbit is one of the most important orbits for commercial spacecraft, and it is also an orbit that can lie outside the magnetopause during large storms. Therefore our analysis will focus on this orbit, a task that is

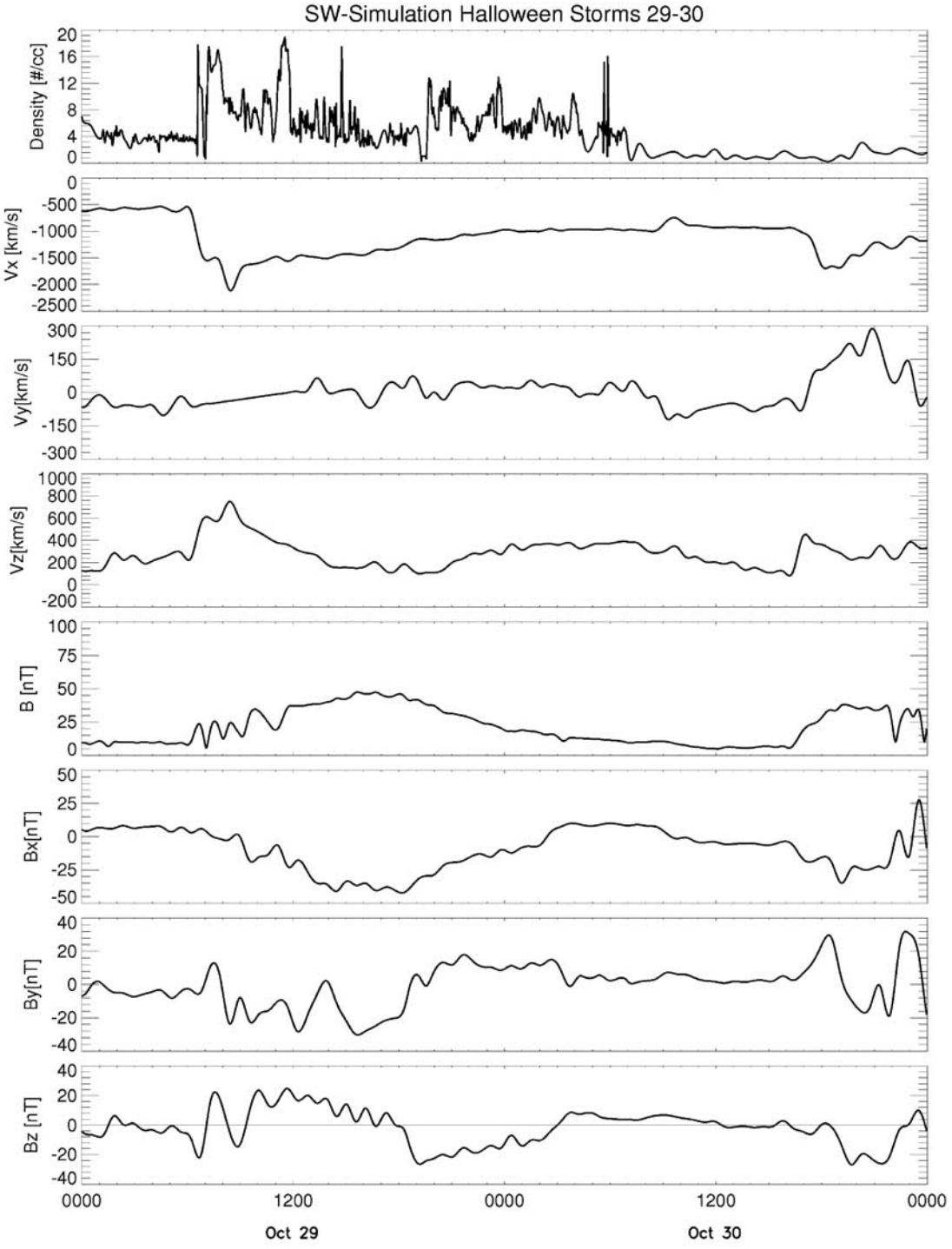

Figure 1. Reconstructed solar wind data as fed into the LFM simulation.

made possible by the fact that geosynchronous satellite data from GOES 10 and 12 are available for this event. Figures 2 and 3 show the geosynchronous magnetic field observations along with the magnetic field as simulated by the LFM interpolated to the positions of the GOES spacecraft. For the days in question GOES 10 was at 135°W longitude (LT = UT − 9) while GOES 12 was at 75°W longitude (LT = UT − 5). The overall agreement in the Z

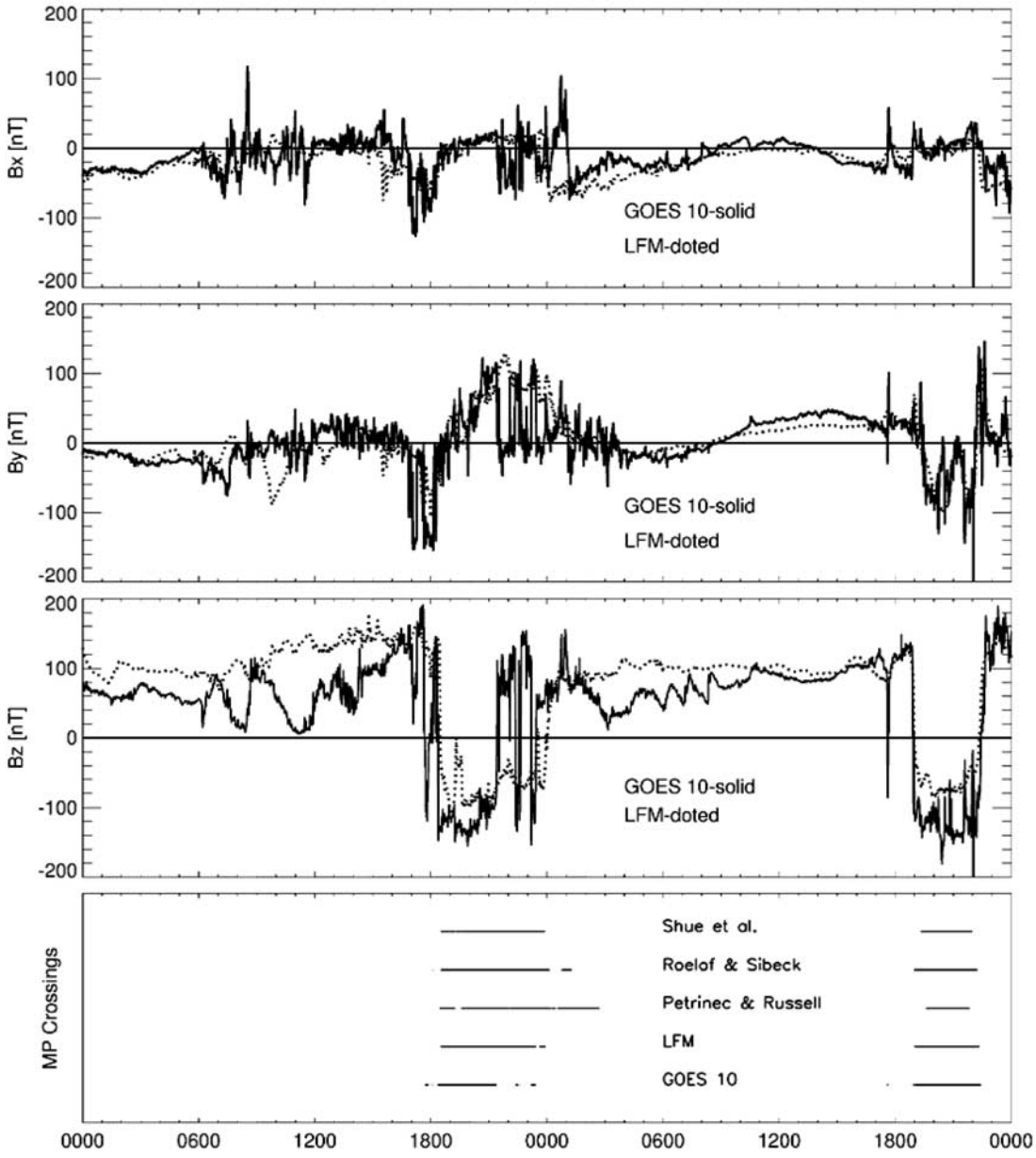

Figure 2. Magnetic field (GSM coordinates) from the GOES 10 spacecraft along with the magnetosphere simulation data interpolated to the satellite position. The bottom plot shows the periods (given by solid lines) when each model (or the actual satellite data) indicated that the satellite was outside the magnetopause.

and Y components is quite good. The correspondence between the simulated and observed X component is not as good as the others, although the overall trends are in the correct direction. The fact that visual inspection shows such a good agreement in general between the simulation and reality leads us to surmise that the reconstructed solar wind data file is not an unreasonable representation of what really hit the Earth's magnetosphere and that the magnetospheric response is also reasonable. However, we want a quantitative measure like a skill score that will allow us to judge the "goodness" of the prediction [e.g., *Murphy*, 1993].

[12] While we might have confidence in the general accuracy of the solar wind data file, given this extreme event, we recognize that we are pushing the empirical models well out of the data parameter space that was used to derive them in the first place. Looking solely at the performance of the empirical models under such solar

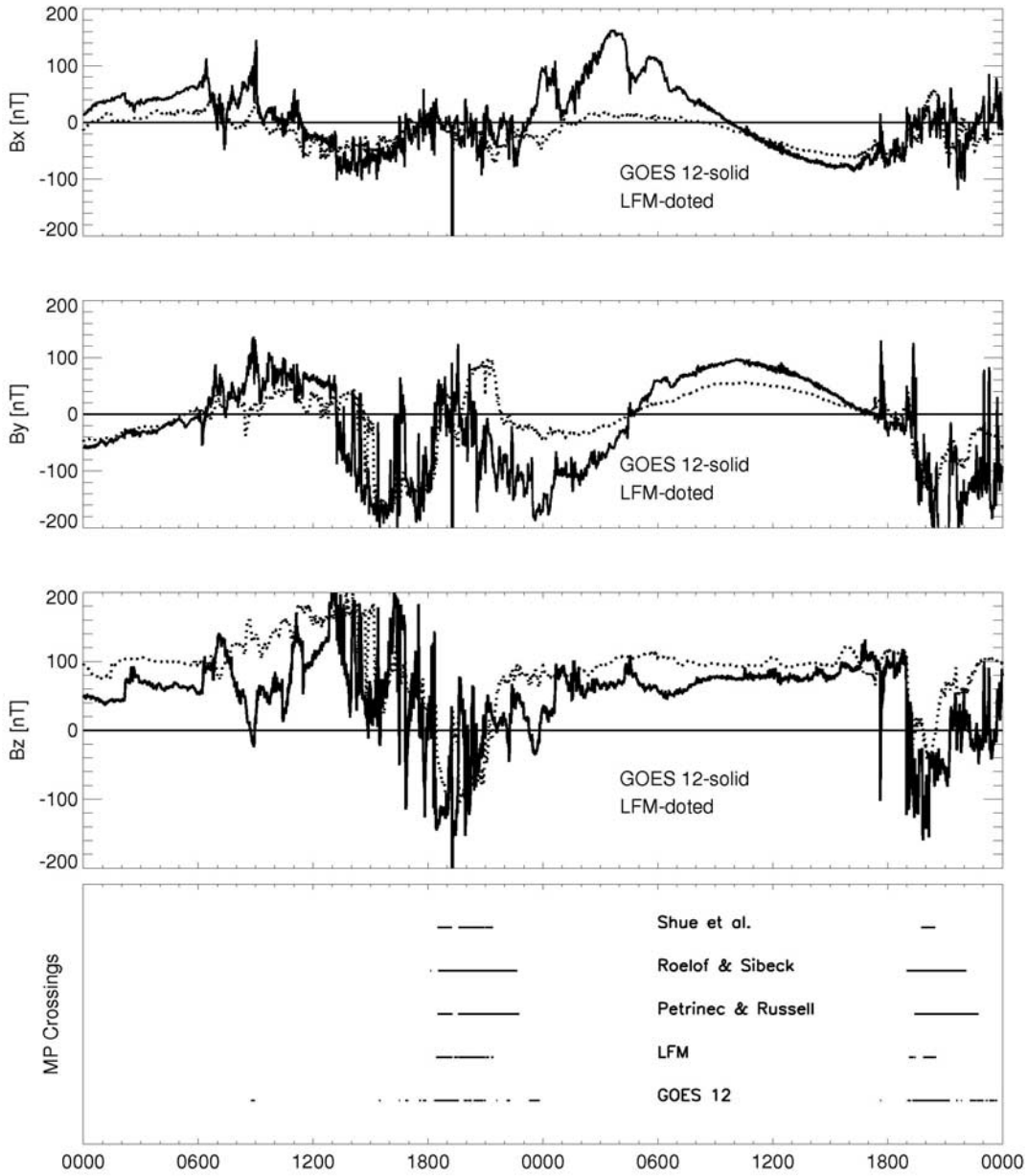

Figure 3. Magnetic field (GSM coordinates) from the GOES 12 spacecraft along with the magnetosphere simulation data interpolated to the satellite position. The bottom plot shows the periods (given by solid lines) when each model (or the actual satellite data) indicated that the satellite was outside the magnetopause.

wind conditions could give a misleading impression of their prediction efficiency. However, when benchmarking the performance of the LFM code in predicting geosynchronous magnetopause crossings we believe that we should begin with the several of the models cited by *Shue et al.* [2000], *Yang et al.* [2002], and *Dmitriev et al.* [2004], despite the fact that we are pushing them out of their valid parameter space. We simply want to provide a quantitative assessment of which models, in an off-the-shelf fashion, can provide the best predictions during the Halloween storms, recognizing that some authors might not want their models so evaluated [*Shue et al.*, 2000].

4. Forecast Verification

[13] The goal of forecast verification is to determine how well a given model is performing. *Murphy* [1993] divided

Table 1. Standard Contingency Table for Dichotomous Forecasts

Forecast/Observations	Yes	No	Total
Yes	H	F	FY
No	M	N	FN
Total	OY	ON	T

the determination of forecast "goodness" into three separate measures. First, a determination is made about how consistent the results match an expert forecaster's best judgment as to what will happen in these circumstances. Second, the quality of how well the forecast matches what actually happened is determined. Finally, the value of a forecast is measured by determining how well it helps a decision maker obtain some benefit. There are numerous aspects to forecast quality. Among them are bias, accuracy, skill, reliability, and resolution. Traditionally accuracy and skill are the leading aspects of determining model quality with the other aspects contributing significantly to the model's value.

[14] The meteorological community has developed an extensive set of tools for measuring the accuracy of predictions, the majority of which are reviewed by *Stanski et al.* [1989]. The simplest category of forecast is for events that have a "Yes" or "No" outcome, e.g., "Will it rain tomorrow?" or "Is a spacecraft outside the magnetopause?" Analysis of these dichotomous forecasts begins with a contingency table, shown in Table 1, which accounts for the four possible combinations of yes/no events for forecasts and observations.

[15] In Table 1, H is the number of hits, F is the number of false alarms, M is the number of misses, and N is the number of correct negatives. A hit represents a forecast event which did occur while a false alarm is a forecast event which did not occur. A miss represents an event which did occur which was not forecast while a correct negative represents no event occurring with a correct forecast. FY is the total number of Yeses forecasts and is the sum of hits and false alarms. The total number of forecast Nos, FN, is the sum of misses and correct negative forecasts. The total number of observed Yeses, OY, is the number of hits plus the number of misses. ON is the total of observed Nos, which is the sum of false alarms and correct negatives. As a final check the sum of FY and FN must equal the sum of ON and OY which is the total number of events in the data set.

[16] We can use the contingency table to calculate a number of different measures that assess the model's ability to forecast correctly. Among these is accuracy (A)

$$A = \frac{H + N}{T}$$

which is a simple measure of the fraction of the correct forecasts. It ranges from 0 to 1 with 1 being a perfect score. It is fairly intuitive to use, but the results can be misleading since it is heavily biased by the most common situation of correct forecasting of No events.

[17] Model bias (B)

$$B = \frac{H + F}{H + M}$$

compares the forecast frequency of Yes events to the observed frequency of Yes events. It ranges from 0 to infinity with 1 being a perfect score. It indicates whether the model has a tendency to under forecast (<1) or over forecast (>1) events. It provides no measurement of how well these forecasts correspond to the observations.

[18] Probability of detection (POD)

$$POD = \frac{H}{H + M}$$

measures the fraction of observed yes events which where correctly forecast. It ranges from 0 to 1 with 1 being a perfect score. This measure is good for rare events, but it can be artificially improved by issuing more yes forecasts to increase hits. It should be used in conjunction with the false alarm ratio (FAR)

$$FAR = \frac{F}{H + F}$$

which measures the fraction of predicted Yes events that did not occur. It ranges from 0 to 1 with 0 being a perfect score. It is sensitive to the climatological frequency of the event.

[19] The probability of false detection (POFD)

$$POFD = \frac{F}{N + F}$$

measures the fraction of No events that were incorrectly forecast as Yes events. It ranges from 0 to 1, with 0 being a perfect score. It is similar to the POD except in this case it can be improved by issuing fewer Yes forecasts and so needs to be used with the POD in order to truly assess the model's capabilities.

[20] In addition to these basic ratios of values in the contingency table a variety of threat scores are used to make quantitative determinations which can be compared between various models for a given interval of interest. The critical success index (CSI), or threat score (TS)

$$CSI = TS = \frac{H}{H + M + F}$$

measures the fraction of observed or forecast events which where correctly predicted. It ranges from 0 to 1 with a perfect score being 1 and 0 indicating no skill. It can be thought of as accuracy with correct No events removed. It

Table 2. Contingency Table for GOES 10 Magnetopause Crossings Predicted by the LFM on 29–30 October[a]

Forecast/Observations	Yes	No	Total
Yes	406	117	523
No	37	2320	2357
Total	443	2437	2880

[a]Values are given as number of minutes.

is sensitive to hits and penalizes both misses and false alarms.

[21] The true skill score (TSS)

$$TSS = POD - POFD = \frac{H}{H+M} - \frac{F}{N+F}$$

measures how well the model separates Yes events from No events. It ranges from −1 to 1, with 0 indicating no skill and 1 being a perfect score. It can be interpreted as (accuracy of events) minus (accuracy of nonevents) minus 1. The true skill score is unduly weighted toward the POD for rare events, so this score is more useful for events that occur frequently. A variant on this skill score is known as the modified true skill score

$$MTSS = \frac{H-M}{H+M} - 2\frac{F}{N}$$

It measures how well the forecast separates the yes events from the no events except in this case it does not over emphasize the POD. It ranges from −1 to 1, with 0 indicating no skill and 1 being a perfect score. The first term is the POD remapped to range of −1 to 1 and the second term penalizes a forecast for predicting a large area for a rare event.

[22] Finally the Heidke skill score

$$HSS = \frac{2(HN - MF)}{(H+M)(M+N) + (H+F)(F+N)}$$

measures the fraction of correct forecasts after eliminating those forecasts that would be correct purely by random chance. It ranges from infinity to 1, with 0 indicating no skill and 1 being a perfect score.

[23] Clearly there are numerous ways to determine the quality of a given forecast. Basic information about the models accuracy and bias will be useful to decision makers in utilizing the results in applications. Another key point is that no single number can be used alone to determine the quality of a given model over any other model.

5. Comparing Model Predictions and Observations

[24] What is of interest from a space weather perspective is if a geosynchronous satellite will cross the magnetopause, and this is a binary prediction; you are either inside or outside. Thus a model does not have to provide an actual magnetopause location. In fact, for the LFM providing the actual magnetopause location might prove difficult since the spatial resolution along the dayside magnetopause is 0.25 R_E or greater, with a consequent smearing of boundaries on at least that scale. However, interpolating the LFM results to the satellite position does allows us to produce a binary prediction, even if we do not know the exact magnetopause position. The bottom plots of Figures 2 and 3 presents this binary information. A solid black line indicates for each model when that model predicts that the satellite will be outside the magnetosphere when the IMF at the Earth is negative. The condition that B_z is negative is essential since we identified magnetopause crossings from the GOES 10 and 12 data using negative B_z as our indicator. The actual times when GOES 10 and 12 were outside the magnetosphere (with negative B_z) are also presented in the same plot.

[25] For each minute on 29 and 30 October, we determined if a particular model indicated that GOES 10 was outside the magnetosphere (as calculated using 1-min resolution solar wind data propagated to the Earth) and the IMF at the Earth had negative B_z. This excludes magnetopause crossings during northward B_z, just as they were excluded from analysis of the GOES data. We recognize that this does skew the comparison somewhat, but magnetopause compressions to, or within, geosynchronous orbit are most often during periods of negative B_z [*Rufenach et al.*, 1989], so we think from a forecasting standpoint, our results are robust. Using this information we are able to calculate the elements of the contingency table for each model. Displaying the contingency table for each model is of little utility, so we only present as a sample the results from LFM for the entire 2-day interval in Table 2. It is interesting to note that GOES 10 only spent 15% of the time outside the magnetopause even under the extreme circumstances of the Halloween storms.

[26] The next step in the forecast verification process is to compute the various measures discussed above. These results are presented in Table 3 for the comparison with GOES 10 and Table 4 for the comparison with GOES 12. In Tables 3 and 4 the metrics are displayed rows with names consistent with the previous formulas. Each model has its

Table 3. Verification Statistics for Each Model Against GOES 10 Observations

	LFM	RS	PR	SA
A	0.95	0.93	0.85	0.93
B	1.18	1.28	1.35	1.09
POD	0.92	0.90	0.72	0.81
FAR	0.22	0.30	0.47	0.25
POFD	0.048	0.069	0.12	0.050
TS	0.73	0.65	0.44	0.64
TSS	0.87	0.83	0.60	0.76
MTSS	0.73	0.64	0.18	0.52
HSS	0.81	0.74	0.53	0.74

Table 4. Verification Statistics for Each Model Against GOES 12 Observations

	LFM	RS	PR	SA
A	0.89	0.88	0.88	0.88
B	0.52	1.02	1.02	0.46
POD	0.40	0.63	0.61	0.34
FAR	0.23	0.39	0.40	0.26
POFD	0.021	0.069	0.073	0.031
TS	0.36	0.45	0.43	0.30
TSS	0.38	0.56	0.54	0.32
MTSS	−0.24	0.10	0.062	−0.37
HSS	0.47	0.55	0.53	0.41

own column with LFM for the Lyon-Fedder-Mobbary model, RS for the model of *Roelof and Sibeck* [1993], PR for the *Petrinec and Russell* [1993] model, and SA for the *Shue et al.* [1998] magnetopause model.

[27] *Shue et al.* [2000] conducted a similar investigation, calculating POD, FAR, and POFD scores for the *Shue et al.* [1998] and *Petrinec and Russell* [1996] models. That study used several years of data with a 20-min separation cadence, with event being a 20-min period when a model predicted that the subsolar standoff distance would be inside of 6.6 R_E at a time when a GOES spacecraft was between 0900 LT and 1500 LT. Thus the *Shue et al.* [2000] study included a number of events less extreme than the Halloween storms that were still able to produce geosynchronous magnetopause crossings. Also, the way in which an ''event'' was defined is different from our ''event'' definition, thus one cannot directly compare our results in Tables 3 and 4 to the results of *Shue et al.* [2000]. *Yang et al.* [2002] also calculated skill scores for magnetopause crossings. While they used a different data set with less extreme events than the Halloween storm, they used the same event definition as we use to develop a contingency table (inside or outside on a 1-min timescale). As a comparison, we note that the POD (0.74) and the FAR (0.27) obtained by *Yang et al.* [2002] for the SA model during the period 1999–2000 are similar to the corresponding values we find in Table 3.

[28] Examining the statistical information in Tables 3 and 4 provides some basic information about how well the model predictions correspond to reality. All the models have an accuracy around 90% which is not surprising since this statistic is heavily influence by correctly forecasting a No event, which is mostly like state for the GOES satellites. It is also clear from these statistics that the models produce better predictions for GOES 10 than for GOES 12. Each model has a higher POD for GOES 10 then for GOES 12 with the FAR and POFD not showing much variation between the spacecraft. All of the models show a reduction in the bias in the GOES 12 data with the LFM and SA models moving from a tendency to over predict crossings to a tendency to under predict them.

[29] The peak POD of 92% was obtained by the LFM for GOES 10 predictions and the worst POD of 34% was obtained by the SA model for GOES 12 predictions. *Shue et al.* [2000] found a much better POD for the SA model, a result that we do not believe is in conflict our result here. The definition of what constitutes an ''event'' is different in our study, and we are in a much different parameter regime during the Halloween storm than the data set used by *Shue et al.* [2000]. Moreover, we are driving all of the models with a ''best guess'' solar wind data file, which will by necessity introduce some error into all of the predictions. The false alarm rate ranged from a low of 22% for LFM and a high of 47% for RS (both for GOES 10). A high false alarm rate is of particular concern in determining the value of a forecast for decision makers. The models that have the highest PODs for the GOES 12 predictions also have the highest FARs.

[30] The threat score removes the prediction of correct negatives from the metric and as such it provides a basic level of assessment of the models abilities. The LFM obtains the highest TS of 73% for GOES 10, but shows a significant reduction to 36% when it comes to predictions for GOES 12. RS obtains a peak value of 65% for GOES 10 and a value of 45% for GOES 12. The PR model shows the least variation between spacecraft for this metric. The SA model has a range comparable to the results from the RS model. Using this statistic we see the LFM obtains the highest score and has an average value for both spacecraft of 55%, which is the same as the average value of the RS model.

[31] While the true skill score measures how well the model separates Yes and No intervals, it is biased toward the POD and is therefore more useful for events that occur frequently, which is not the case for geosynchronous magnetopause crossings. On the other hand, the modified true skill score penalizes models for predicting large Yes intervals for rare events and thus is well suited to the assessment of magnetopause crossings. Once again the LFM obtains the highest value for this metric for its predictions of GOES 10. It is also clear from this metric that none of the models is accurately predicting GOES 12 crossings because a value of 0 indicates that models have no skill and only RS and PR obtain values slightly above 0 for this comparison.

[32] The predictions for GOES 12 seem to be not as good as the predictions for GOES 10, with the exception of PR, where the overall predictions were comparable for GOES 12. This result is not likely that this is due to major errors in the solar wind data file, because the GOES 10 predictions at the same time are much better. In addition, the general behavior of the magnetic field at GOES 12 in the LFM simulation is similar to that recorded by GOES 12, it is just that the number of negative B_z intervals predicted by LFM are less, especially on 30 October. The only difference is that GOES 12 leads GOES 10 by four hours in local time.

[33] *Dmitriev et al.* [2005] came to the same conclusion; there is a pronounced asymmetry between GOES 10 and 12 during the Halloween storm. Moreover, *Dmitriev et al.*

[2004] presented clear statistical evidence of a local time asymmetry in geosynchronous magnetopause crossings. They posited two possible explanations. One was the effect of the asymmetric ring current during the main phase of a storm, and the other is asymmetric magnetopause erosion. Given the extreme nature of the Halloween storms, it is not surprising that the local time asymmetry is so pronounced, since either mechanism would be exaggerated in this case.

[34] We do not consider asymmetric erosion to be the most likely origin of the asymmetry. The LFM code does a good job of capturing the large-scale physics of magnetopause erosion [*Wiltberger et al.*, 2003], so if a global aspect of solar wind–magnetosphere coupling like erosion were the origin of the asymmetry, we would expect to capture it in the simulation. On the other hand, the ring current explanation is certainly consistent with the fact that the LFM does not include the energy-dependent drifts in the inner magnetosphere. If the ring current is indeed the origin of the asymmetry, once the LFM is coupled to the Rice Convection Model [e.g., *Goodrich et al.*, 2004], which will provide a ring current with multiple species and energy-dependent drifts, we should see an increase in the skill scores for GOES 12 during this event. If we see no improvement in the skill scores once a ring current is added one could surmise that additional physics not included in the model is the origin of the asymmetry. In general, there should increase in skill scores over time as increasingly sophisticated models become available, as happened in terrestrial meteorology [*Siscoe*, 2006].

[35] One further point concerns the reason (or reasons) why we get the results we get, for all of the models. Although we pushed the empirical models well outside of their bounds of validity [e.g., *Shue et al.*, 2000], they actually performed rather well. In fact, it was not clear at the outset that the LFM would be a better predictor of geosynchronous magnetopause crossings than the empirical models, and actually the differences in the skill scores are not overwhelming. However, the LFM is able to handle a wide range of solar wind conditions from first principles. The LFM simulation code, being dynamic, could also represent in a more realistic fashion the response of the magnetospheric boundary to the solar wind variations, as opposed to models that are self-similar analytic solutions that cannot represent local boundary motions. One might also want to ask why a particular empirical model scored better than another. Such a discussion is beyond the scope of this paper, which focuses solely on a quantitative assessment of the prediction efficiency of geosynchronous magnetopause crossings during an extreme event. For the purpose of this study is enough to know that the LFM code does as well (or even a bit better) at predicting geosynchronous magnetopause crossings during the Halloween storms and to document the skills scores. This extreme event is a period that we will use as benchmark to determine the relative performance of improved CISM models to determine if we are indeed moving up the skill curve, especially with regard to local time asymmetry.

6. Conclusions

[36] We have assessed the ability of several models to predict whether the GOES spacecraft will be in or out of the magnetosphere during the Halloween storms of 2003, with the goal of providing a benchmark against which to measure the skill of CISM simulations in the future. The Lyon-Fedder-Mobarry simulation had the best overall set of skill scores, though these scores were not dramatically better that the empirical models. We conclude that the LFM provides the best prediction for magnetopause crossing for the extreme conditions present during the Halloween storm. This result is not entirely surprising since the empirical models were pushed well beyond the range of validity because of the extreme solar wind conditions, while the LFM code is relying on basics physics to make its predictions and thus can handle extreme conditions as long as the physics is not altered by those conditions. It should be noted that the most valuable space weather predictions are likely to be during extreme conditions, when the effects of space weather are greatest.

[37] Following previous results [*Dmitriev et al.*, 2004, 2005], we find a significant local time asymmetry in the magnetopause position. While it is unclear exactly what is the cause of this asymmetry, it should be possible to determine if the asymmetry is ring current related using future CISM magnetospheric models that incorporate the ring current. These results will allow us to benchmark future improvements in the LFM code, and well as additional codes that comprise LFM coupled to a ring current code. However, all of the comparisons have been done for southward IMF so that we could easily identify crossings. We do not know if the skills scores we developed are representative of northward IMF magnetopause crossings, which should be rare in any case. Nonetheless, it seems reasonable to develop a robust method for identifying such crossings in the simulation.

[38] **Acknowledgments.** The authors would like to thank Terry Onsager and Howard Singer for useful conversations concerning space weather metrics of interest to the Space Environment Center. This material is based upon work supported by CISM, which is funded by the STC Program of the National Science Foundation under agreement ATM-0120950, and NASA grant NAG5-1057.

References

Carlowicz, M. J., and R. E. Lopez (2002), *Storms from the Sun*, Joseph Henry Press, Washington, D. C.

Dmitriev, A. V., A. V. Suvorova, J. K. Chao, and Y.-H. Yang (2004), Dawn-dusk asymmetry of geosynchronous magnetopause crossings, *J. Geophys. Res.*, *109*, A05203, doi:10.1029/2003JA010171.

Dmitriev, A., J.-K. Chao, M. Thomsen, and A. Suvorova (2005), Geosynchronous magnetopause crossings on 29–31 October 2003, *J. Geophys. Res.*, *110*, A08209, doi:10.1029/2004JA010582.

Elkington, S. R., M. Wiltberger, A. A. Chan, and D. N. Baker (2004), Physical models of the geospace radiation environment, *J. Atmos. Sol. Terr. Phys.*, *66*, 1371–1387, doi:10.1016/j.jastp.2004.03.023.

Goodrich, C. C., A. L. Sussman, J. G. Lyon, M. A. Shay, and P. A. Cassak (2004), The CISM code coupling strategy, *J. Atmos. Sol. Terr. Phys.*, *66*, 1469–1479, doi:10.1016/j.jastp.2004.04.010.

Hughes, W. J., and M. K. Hudson (2004), Towards an integrated model of the space weather system, *J. Atmos. Sol. Terr. Phys.*, *66*, 1241, doi:10.1016/j.jastp.2004.06.001.

Lanzerotti, L. J. (2003), Space weather's time has come, *Space Weather*, *1*(1), 1003, doi:10.1029/2003SW000024.

Lopez, R. E., C. Goodrich, M. Wiltberger, and J. Lyon (2000), Solar wind-magnetosphere energy coupling under extreme interplanetary conditions: MHD simulations, *J. Atmos. Sol. Terr. Phys.*, *62*, 865–874.

Lopez, R. E., D. Baker, and J. Allen (2004), Sun unleashes Halloween storm, *Eos Trans. AGU*, *85*, 105.

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, doi:10.1016/j.jastp.2004.03.020.

Murphy, A. H. (1993), What is a good forecast? An essay on the nature of goodness in weather forecasting, *Weather Forecasting*, *8*, 281–293.

Onsager, T. G., B. Poppe, and W. Murtagh (2004), Halloween storms star in Space Weather Week 2004 Meeting, *Space Weather*, *2*(8), S08003, doi:10.1029/2004SW000099.

Petrinec, S. M., and C. T. Russell (1993), An empirical model of the size and shape of the near-earth magnetotail, *Geophys. Res. Lett.*, *20*(23), 2695–2698.

Petrinec, S. M., and C. T. Russell (1996), Near-Earth magnetotail shape and size as determined from the magnetopause flaring angle, *J. Geophys. Res.*, *101*(A1), 137–152.

Roelof, E. C., and D. G. Sibeck (1993), Magnetopause shape as a bivariate function of interplanetary magnetic field B_z and solar wind dynamic pressure, *J. Geophys. Res.*, *98*(A12), 21,421–21,450.

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

Shue, J.-H., et al. (1998), Magnetopause location under extreme solar wind conditions, *J. Geophys. Res.*, *103*(A8), 17,000–17,691.

Shue, J.-H., J. K. Chao, H. C. Fu, C. T. Russell, P. Song, K. K. Khurana, and H. J. Singer (1997), A new functional form to study the solar wind control of the magnetopause size and shape, *J. Geophys. Res.*, *102*(A5), 9497–9512.

Shue, J.-H., P. Song, C. T. Russell, J. K. Chao, and Y.-H. Yang (2000), Toward predicting the position of the magnetopause within geosynchronous orbit, *J. Geophys. Res.*, *105*(A2), 2641–2656.

Siscoe, G. (2006), A culture of improving forecasts: Lessons from meteorology, *Space Weather*, *4*, S01003, doi:10.1029/2005SW000178.

Siscoe, G. L., W. J. Hughes, H. E. Spence, D. N. Baker, and R. S. Weigel (2004), Roles of empirical modeling with CISM, *J. Atmos. Sol. Terr. Phys.*, *66*, 1481–1489, doi:10.1016/j.jastp.2004.03.028.

Spence, H., D. Baker, A. Burns, T. Guild, C.-L. Huang, G. Siscoe, and R. Weigel (2004), Center for integrated space weather modeling metrics plan and initial model validation results, *J. Atmos. Sol. Terr. Phys.*, *66*, 1499–1507, doi:10.1016/j.jastp.2004.03.029.

Stanski, H. R., L. J. Wilson, and W. R. Burrows (1989), Survey of common verification methods in meteorology, *World Weather Watch Tech. Rep. 8*, *WMO/TD 358*, 114 pp., World Meteorol. Org., Geneva, Switzerland.

Wiltberger, M., R. E. Lopez, and J. G. Lyon (2003), Magnetopause erosion: A global view from MHD simulation, *J. Geophys. Res.*, *108*(A6), 1235, doi:10.1029/2002JA009564.

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, doi:10.1016/j.jastp.2004.03.026.

Yang, Y.-H., J. K. Chao, C.-H. Lin, J.-H. Shue, X.-Y. Wang, P. Song, C. T. Russell, R. P. Lepping, and A. J. Lazarus (2002), Comparison of three magnetopause prediction models under extreme solar wind conditions, *J. Geophys. Res.*, *107*(A1), 1008, doi:10.1029/2001JA000079.

C. C. Goodrich, C.-L. Huang, E. L. Kepko, and H. Spence, Department of Astronomy, Boston University, Boston, MA 02215, USA.

S. Hernandez and R. E. Lopez, Department of Physics and Space Sciences, Florida Institute of Technology, Melbourne, FL 32901, USA. (relopez@fit.edu)

J. G. Lyon, Department of Physics and Astronomy, Dartmouth College, Hanover, NH 03755, USA.

M. Wiltberger, HAO, NCAR, 3080 Center Green Dr, Boulder, CO 80301, USA.