AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Space Physics

RESEARCH ARTICLE
10.1002/2015JA021825

Key Points:
- A robust algorithm is developed to detect and extract regions 1 and 2 field-aligned current features
- Increased resolution of LFM model improves FAC simulation and reduces bias against AMPERE estimates
- The introduction of a statistical framework allows for quantitative estimates of model discrepancy

Correspondence to:
W. Kleiber,
william.kleiber@colorado.edu

Citation:
Kleiber, W., B. Hendershott, S. R. Sain, and M. Wiltberger (2016), Feature-based validation of the Lyon-Fedder-Mobarry magnetohydrodynamical model, *J. Geophys. Res. Space Physics*, *121*, 1192–1200, doi:10.1002/2015JA021825.

Received 18 AUG 2015
Accepted 15 JAN 2016
Accepted article online 19 JAN 2016
Published online 6 FEB 2016

Feature-based validation of the Lyon-Fedder-Mobarry magnetohydrodynamical model

W. Kleiber[1], B. Hendershott[1], S. R. Sain[2], and M. Wiltberger[3]

[1]Department of Applied Mathematics, University of Colorado, Boulder, Colorado, USA, [2]Climate Corporation, San Francisco, California, USA, [3]High Altitude Observatory, National Center for Atmospheric Research, Boulder, Colorado, USA

Abstract Field-aligned currents (FACs) play an important role in the coupling between the ionosphere and magnetosphere. Numerical simulation of these phenomena is of increasing interest, but validation has been hampered by a lack of a formal framework to compare simulations to satellite-derived products. We develop a statistical approach to compare FAC simulations from global magnetohydrodynamical models against satellite products. We introduce a robust algorithm that automatically detects and defines regions 1 and 2 FACs. In an example, currents derived from the Iridium satellites are compared against simulated currents from two resolutions of the Lyon-Fedder-Mobarry model on one solar event. We assess both average and structured discrepancies, the former being a level shift of the physical model away from the satellite product, while structural discrepancy refers to time-varying, continuous differences. For this event, the lower resolution version of the Lyon-Fedder-Mobarry is shown to be a poor representation of the satellite-derived FACs, while the higher resolution version substantially reduces discrepancy.

1. Introduction

A critical difficulty in developing geophysical models is validation; that is, assessing to what extent the model is an adequate representation of reality. Typically, model validation is performed in an ad hoc way that involves expert understanding and interpretation of model solutions, which sometimes can take the form of a quantitative assessment. The aim of this paper is to develop a validation framework for the global magnetospheric models of the near-Earth space environment with particular focus on field-aligned currents (FACs). We illustrate the method by validating the Lyon-Fedder-Mobarry (LFM) model against a set of Active Magnetosphere and Polar Electrodynamics Response Experiment (AMPERE)-derived FACs.

We propose a statistical approach to quantify and account for relevant sources of uncertainty when performing model validation, and that allows for formal testing of scientific hypotheses. Briefly, the approach relies on developing a feature detection algorithm specifically designed to extract regions 1 and 2 currents from the field-aligned current (FAC) simulation in the magnetospheric model as well as observational or satellite-derived data products, followed by a statistical comparison of the two sets of currents. The feature detection algorithm developed here works well with the FAC comparison but could be adapted to other spatial quantities, e.g., auroral images or solar magnetograms.

Field-aligned currents, also known as Birkeland currents, between the magnetosphere and ionosphere represent one of the fundamental mechanisms for coupling these regions. Using observations from the Triad satellite, *Iijima and Potemra* [1978] constructed a map of large-scale structure of these current systems. In the same work they also introduced concepts of regions 1 and 2 FAC patterns to describe the large-scale features present in the observations during substorm intervals. These current patterns are also produced along with the ionospheric convection and conductances using ground- and space-based observations by the assimilative mapping of ionospheric electrodynamics technique [*Richmond and Kamide*, 1988; *Richmond et al.*, 1998]. FAC patterns can also be derived statistically [*Weimer*, 2000, 2001].

More recently, the temporal evolution of the FAC structure has been observed by the Active Magnetosphere and Polar Electrodynamics Response Experiment (AMPERE) [*Anderson et al.*, 2002]. These studies involve using the engineering magnetometer data on board the 66+ spacecraft distributed in six equally spaced polar 780 km altitude orbit planes. The contraction of the FAC patterns involves a spherical harmonic fit to the observations collected over a time period of approximately 10 min. The details of this process and the limitations of

data reconstruction are described by *Waters et al.* [2001]. A data product routinely produced from this process is a pattern of FACs on a polar magnetic grid with 15° resolution longitude and 1° resolution in latitude.

Numerous models now exist to simulate the interaction of the solar wind with the magnetosphere-ionosphere system [*Lyon et al.*, 2004; *Wiltberger et al.*, 2004; *Raeder et al.*, 2008; *Tóth et al.*, 2012; *Winglee et al.*, 1998; *Janhunen*, 1996]. While there are significant differences in the numerical methods and capabilities of these models, each includes a simulation of the electrodynamic coupling between the ionosphere and magnetosphere. *Ridley et al.* [2010] examined the impact of resolution, numerical scheme, and other parameters on the ionospheric coupling, including the cross polar cap potential pattern and field-aligned currents within the Space Weather Modeling Framework. *Merkin and Lyon* [2010] reported on how the choice of low-latitude boundary condition on the electrodynamic solution can impact the structure of the magnetosphere-ionosphere system. *Raeder and McPherron* [1998] examined the structure of the substorm current wedge within the OpenGGCM framework. Validation studies conducted by the Community Coordinated Modeling Center compared the model's ability to reproduce geosynchronous magnetic field observations [*Rastaetter et al.*, 2011], the D_{st} index [*Rastaetter et al.*, 2013], and ground magnetic field perturbations [*Pulkkinen et al.*, 2011, 2013]. Additionally, the LFM model has been a subject of statistical calibration and emulation based on approximated features [*Kleiber et al.*, 2013, 2014; *Heaton et al.*, 2015]. It is noteworthy that these community, metric-based validation studies have focused on comparisons with time series data sets at a collection of observations and not with spatially distributed observational patterns.

Comparisons between the Iridium FAC reconstructions and results of global simulations are of prime interest. *Korth et al.* [2004] compared the observations of FAC patterns during periods with steady solar wind conditions for several hours to those produced with the LFM simulation. They found that the low resolution versions of the simulations used did not reproduce region 2 FACs. Using an ad hoc detection method, *Korth et al.* [2008] found that region 1 FACs in the LFM were approximately 5° poleward of the location seen in the AMPERE data sets. More recently, *Merkin et al.* [2013] used high-time resolution AMPERE data along with high-resolution LFM simulations to qualitatively show that the simulation reproduced the global features of the 2-D FAC pattern. We note that while we illustrate our method on the LFM model, other models may also be entertained, such as OpenGGCM [*Vennerstrom et al.*, 2005], Grand Unified Magnetosphere-Ionosphere Coupling Simulation (GUMICS) [*Juusola et al.*, 2014], or the Tanaka model [*Tanaka*, 2000].

2. Event and Simulation Data

The magnetospheric event we examine in this paper occurred on 3 August 2010 over approximately 90 min over the Northern Hemisphere. The AMPERE product is from the Iridium satellite constellation, and this particular event was previously examined by *Merkin et al.* [2013], who discuss the FAC derivation scheme. The AMPERE product contains 2 min increments consisting of 10 min averaged values over 1707 UT to 1833 UT, resulting in FAC fields at 44 time points. Within this time frame, an interplanetary shock arrived at approximately 1740 UT, followed by a coronal mass ejection.

The version of the LFM we use in this paper is run on massively parallelized super computers, and whose simulations are available at multiple resolutions [*Lyon et al.*, 2004]. Below we use two resolutions, which we will refer to as *double* and *octuple*. The double resolution is the current standard implementation, while the octuple resolution quadruples the resolution in each dimension. In particular, the double resolution has a grid of dimension $53\times48\times64$, while the octuple resolution has a grid of dimension $212\times192\times256$ in a nonuniform distorted spherical grid (see *Merkin et al.* [2013] for details). The lower resolution version is computationally feasible and can be run in almost real time, while the octuple version requires substantially more time on a super computer.

Basic hypotheses regarding these data and simulations are that the octuple resolution provides a better representation of the underlying physical process than the double resolution version of the LFM and, additionally, that the octuple resolution version is an adequate representation of the process. In the next section we describe an approach to extract relevant FAC features from these various data sources, and the ensuing sections provide a statistical framework for formally assessing differences between these differing model resolutions and the AMPERE product.

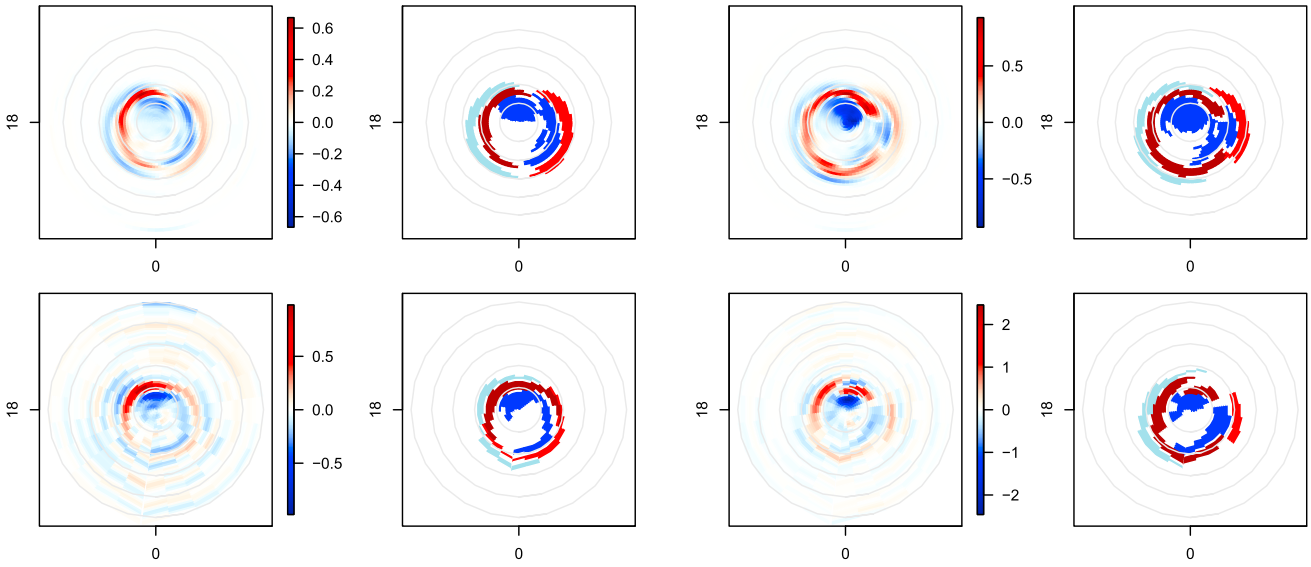

Figure 1. (bottom row) AMPERE product and (top row) octuple LFM with associated detected features at 1707 UT (columns 1 and 2) and 1825 UT (columns 3 and 4). Dark red (blue) corresponds to region 1 positive (negative) feature, while light red (blue) corresponds to region 2 positive (negative) feature. Color bar units are $\mu A/m^2$, and coordinates are in magnetic local time thermosphere; grey circles are 10° increments of latitude from the pole.

3. Feature Detection

Our approach to statistical validation of the LFM begins by applying a robust feature detection algorithm to both the LFM and AMPERE product with a goal of identifying regions 1 and 2 positive and negative current features that can then be directly compared. We give a brief overview of our detection algorithm here, leaving the technical details to Appendix A.

The feature detection algorithm has three main components. Initially, the data are spatially smoothed to help make features more pronounced and reduce noise. Second, the spatial field at each time point of the event is segmented into a set of potential spatial features by thresholding. Finally, regions 1 and 2 features are chosen from the initial segments using an algorithm that favors features exhibiting temporal persistence as well as nightside and dayside symmetries.

Figure 1 illustrates some example features detected by our algorithm at two different time points for LFM and AMPERE spatial fields. The algorithm successfully identifies and separates the salient regions 1 and 2 positive and negative features present in the LFM model. We see similar performance for the AMPERE product, which exhibits a more challenging scenario with substantial ringing artifacts in the data. Nonetheless, our algorithm identifies features that correspond to and may be directly compared with the LFM features. Note that we are able to identify features that contain spatially disjoint regions, such as region 1 negative feature at 1825 UT in the AMPERE product. A simple thresholding approach to feature identification would be unable to capture such groupings. We note that the algorithm would need to be slightly adapted if higher-latitude region 0 currents were present.

4. Validation and Analysis

We now describe a set of statistical procedures to assess the adequacy of the LFM's ability to replicate the AMPERE-derived estimate of FACs. Although we expect the higher resolution run to produce more accurate representations of the AMPERE product than the lower resolution run, it is crucial to quantify the difference between resolutions.

Two important physical components are the space-time extents of positive and negative regions 1 and 2 currents and the respective magnitudes of currents flowing through these regions. For each validation time point $t = 1, \ldots, 44$ (spanning 1707 UT to 1833 UT) of the event we statistically model the integrated current of the AMPERE product, Y, double resolution LFM, D, and octuple resolution LFM, O, as

Table 1. Estimates of the Level Discrepancy Shift of the Total Integrated Current Between the Double (*D*) and Octuple (*O*) Resolution LFM Runs as Compared To the AMPERE Product[a]

$\mu_{D,1,+}$	$\mu_{O,1,+}$	$\mu_{D,1,-}$	$\mu_{O,1,-}$
−0.39 (0.024)	−0.07 (0.030)	0.80 (0.049)	−0.10 (0.034)

[a]Biases are shown for region 1 positive (+) and negative (−) currents. The octuple resolution version reduces bias by between 82% and 87% over the double resolution. Units are log μA/m^2.

$$\begin{aligned}\log|Y_{R,C}(t)| &= \mu_{R,C} + S_{R,C}(t) + \varepsilon_{R,C}(t)\\ \log|D_{R,C}(t)| &= \mu_{R,C} + \mu_{D,R,C} + S_{R,C}(t) + \delta_{D,R,C}(t) + \varepsilon_{D,R,C}(t)\\ \log|O_{R,C}(t)| &= \mu_{R,C} + \mu_{O,R,C} + S_{R,C}(t) + \delta_{O,R,C}(t) + \varepsilon_{O,R,C}(t)\end{aligned} \tag{1}$$

separately for each region $R = 1, 2$ and current $C =$ positive, negative.

The AMPERE-derived measurements are decomposed as a steady underlying mean state $\mu_{R,C}$ plus a dynamical term that evolves over time, $S_{R,C}(t)$. The simulation runs may exhibit some discrepancy from the satellite product. The introduction of the logarithm is a standard statistical practice to reduce skewness and bring the model residuals closer to normality; we note that the effect of this is that model errors are multiplicative, which is reasonable for variables that are unsigned.

The process model contains a level mean $\mu_{R,C}$ that is present for both the AMPERE product and model runs, while the model runs contain additional mean shifts $\mu_{D,R,C}$ or $\mu_{O,R,C}$ that allow for possible overall magnitude errors of the LFM simulation. To capture the temporally evolving structure of the process over time, we include a dynamic mean curve $S_{R,C}(t)$ that is parameterized as a centered B spline with four knots across the time domain. The double and octuple resolution LFM runs include the crucial $\delta_{D,R,C}(t)$ and $\delta_{O,R,C}(t)$ quantities, respectively, that are dynamical discrepancy functions. These discrepancy functions are a critical quantity of interest and represent the amount by which (up to constant shifts in magnitude) the LFM adequately represents the AMPERE product. If the discrepancies are estimated as $\delta \equiv 0$, the LFM simulation adequately captures important physical aspects of the AMPERE product without apparent bias. Finally, the error processes $\varepsilon_{R,C}$, $\varepsilon_{D,R,C}$, and $\varepsilon_{O,R,C}$ are assumed to be Gaussian white noise processes that are mutually independent.

Parameters of the statistical model (1) are estimated by ordinary least squares. Q-Q plots suggest that the normality assumption of the errors is justifiable (not shown). The estimated error processes exhibit some residual temporal correlation; we experimented with more sophisticated statistical models that capture this correlation; however, the results were quantitatively similar, so we maintain the current model formulation. Our model distinguishes between *average* bias and *structured* bias. Average bias refers to the $\mu_{\cdot,\cdot,\cdot}$ term, whereas structured bias refers to the time-evolving term $\delta_{\cdot,\cdot,\cdot}(t)$. Table 1 shows the estimates and standard errors for the average bias for the positive and negative region 1 features. Both the double and octuple resolutions have statistically significant average bias shifts from the AMPERE product, indicating that possibly higher resolutions than octuple are necessary to adequately reproduce the satellite estimates or that there are biases present in the satellite product. However, the octuple version reduces bias by between 82% and 87% over the double resolution version, resulting in about an order of magnitude reduction of bias. Thus, there is evidence that the increased resolution version improves the LFM fit of the AMPERE product at an average level given the dynamical terms.

To assess the structured bias terms, Figure 2 shows the time series for integrated current in region 1 positive and negative features. Heuristically, the octuple resolution runs of LFM appear to represent the AMPERE product substantially more so than the double resolution LFM. To formally quantify the visual difference, Figure 2 (bottom row) shows estimates of the structured bias term $\hat{\delta}_{\cdot,1,\cdot}(t)$ along with 95% joint confidence intervals. The double resolution LFM displays relatively large deviations from zero in the discrepancy functions, while the octuple resolution version generally contains zero in the uncertainty bounds. Echoing the results for the level bias term, the octuple resolution version of the LFM is not statistically unbiased; however, there is substantial evidence that the structural discrepancy from the AMPERE product is greatly reduced by using the higher resolution version and indeed produces an almost unbiased replication of the satellite estimate. Remaining error may be due to effects outside the realm of the MHD simulation.

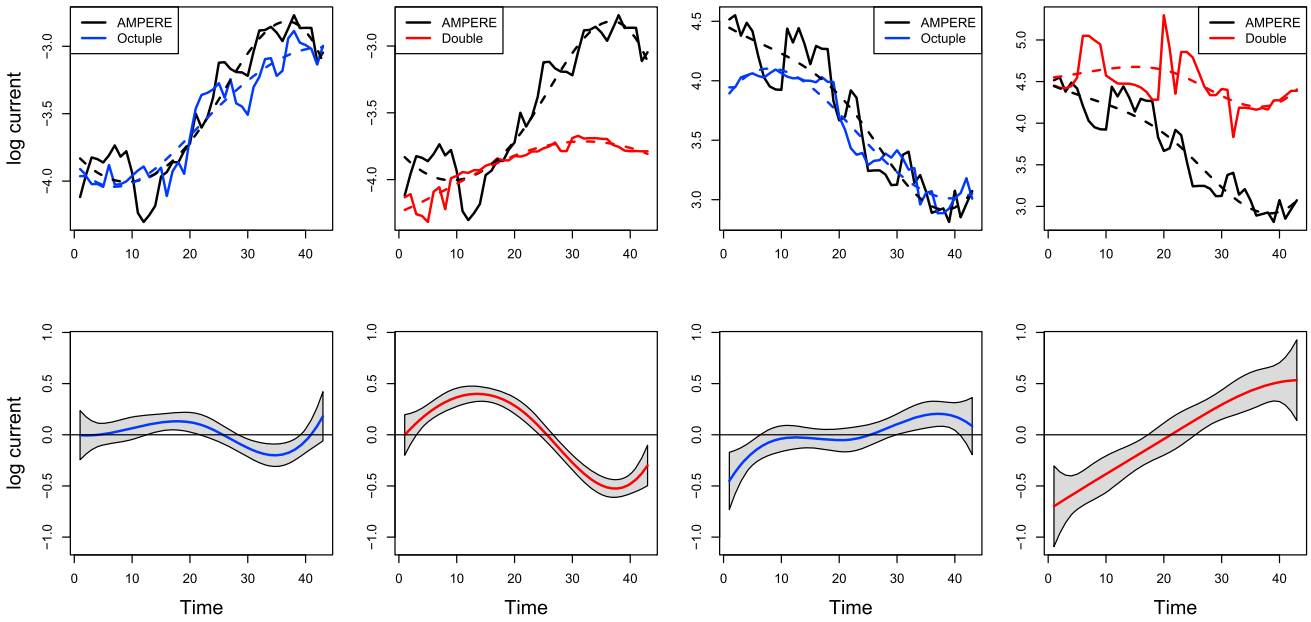

Figure 2. (top row) Comparison of octuple and double resolution LFM with the AMPERE region 1 features. Solid lines are spatially integrated currents, and dashed lines are the fitted mean current curve for the octuple resolution (blue), double resolution (red) LFM runs, and AMPERE product (black). Curves are shown for region 1 positive feature (columns 1 and 2) and negative feature (columns 3 and 4). (bottom row) Estimated discrepancies $\delta_{\cdot,R,C}(t)$ with joint 95% confidence intervals. Units are log μA/m^2.

The next step in our analysis considers the structure of the identified features, comparing the total spatial areas of regions 1 and 2 features between the octuple resolution LFM and AMPERE product. The double resolution LFM has many cases where region 2 currents are not present in the simulation, echoing *Korth et al.* [2004], and given the relatively poor performance seen in Figure 2, we remove it from consideration. Additionally, exploratory analysis suggested that inclusion of the logarithm terms of (1) was not crucial, so it is removed from the statistical model. Areas considered are based on a unitless equal area azimuthal projection in which parallels are concentric circles and meridians are equally spaced radial lines.

Figure 3 shows the temporal evolution of regions 1 and 2 feature areas, respectively, for the octuple resolution LFM and AMPERE product. There is clear evidence of a level shift between the model and observed features, in that for both regions 1 and 2 positive and negative features, the LFM simulates features that are inflated compared to AMPERE. The associated discrepancy functions are also shown, and generally center about zero, with some evidence of spatially varying bias. Region 1 negative feature discrepancy is not statistically different than zero, suggesting that the LFM has a dynamical bias for this feature. Note that the magnitudes of the discrepancy functions are approximately one order less than that of the level shift ($\mu_{O,\cdot,\cdot}$), suggesting that we can attribute most structural problems with the LFM to a constant bias in regions 1 and 2 feature areas.

5. Summary and Future Directions

Validation of global simulations of the magnetosphere-ionosphere system against observational data is a crucial and difficult endeavor. A statistical approach is preferred, as it allows for formal quantification of various sources of uncertainty and quantitative statements regarding adequacy of model fit (e.g., adequacy of the LFM representation of the AMPERE product). We introduced a feature detection algorithm specifically designed for robust field-aligned current detection and a statistical approach to test and assess the quality of the LFM model in replicating satellite-derived FACs during a geomagnetic disturbance. We applied the feature detection algorithm to the AMPERE product and the LFM at two resolutions and compared the total integrated currents and spatial extents of these features between the model and satellite estimates of FACs.

Our results support the scientific assessment that the higher resolution run is a better representation of the satellite-estimated process than the lower resolution version. Indeed, the lower resolution version is widely unable to generate region 2 currents throughout the geomagnetic disturbance. However, we also found that there are still some statistically significant discrepancies between the LFM and AMPERE product even at a

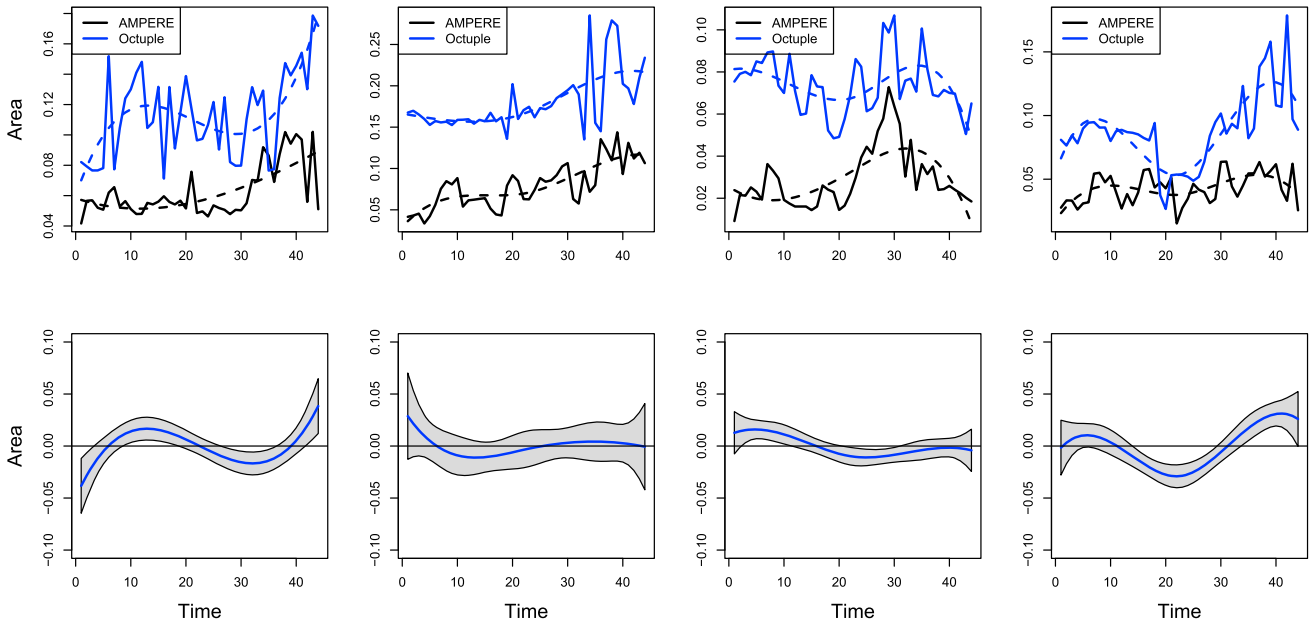

Figure 3. (top row) Areas of region 1 positive and negative features (columns 1 and 2, respectively) and areas of region 2 positive and negative features (columns 3 and 4, respectively) for the octuple resolution LFM and AMPERE product. (bottom row) Estimated discrepancy function $\delta_{\cdot,\cdot,\cdot}(t)$ with 95% joint confidence intervals.

fine resolution, some of which can be attributed to model errors but the remainder of which indicates some structural discrepancies between LFM's space-time evolution of the field versus that present in the AMPERE estimate.

Future research may be devoted at adapting our algorithm to other quantities and considering other geomagnetic disturbance events. Indeed, it is a priori unclear that any LFM discrepancies will be consistent across events, and further research is warranted to assess such a possibility. Statistical research may be directed toward capturing the uncertainty in the algorithmic output, which can then be integrated into a more complete understanding of the various sources of uncertainty and error in the physical and statistical models.

Appendix A: Feature Detection Algorithm

The feature detection algorithm has three main components. Initially, the data are smoothed to help make features more pronounced and reduce noise; second, each time point is segmented into a set of potential spatial features; and finally, regions 1 and 2 features are chosen from the initial set. The following subsections detail each aspect of the full feature detection algorithm.

A1. Smoothing

In smoothing the data, we utilize a separable Gaussian kernel that separately weights the radial and angular directions. The kernel is the product of two functions $K_\lambda(r) = \exp(-(r/\lambda)^2)$, r indexing either latitude or angular distance and λ (degrees latitude) acting as a bandwidth parameter that controls the amount of smoothing. We set $\lambda = 0.0005$ and 0.25 for latitude and angle, respectively, using Euclidean distance over latitude and angular distance over angle. The results of the smoothing are twofold: the prevalence of ringing effects and noise in the AMPERE product and model data are reduced, and the coherence of features in the model output is increased considerably. The smoothing is performed on the spatial fields separately for each time point.

A2. Image Segmentation

After smoothing the data, we proceed to determine which parts of the data are features that might be included in region 1 or 2 currents. To this end, a novel image segmentation algorithm is required. The algorithm, which is applied separately to each spatial field at all time points, takes the following steps:

1. Find all local minimums and maximums of the data.
2. Starting at the largest local maximum, add this point to the current segment then check the neighboring data points of this starting location to see if they are above a certain threshold (we use the 88% quantile

based on positive values for the positive features and the 24% quantile based on negative values for the negative features). These values were identified after extensive exploratory testing.
3. If a neighboring point is above the threshold, it is added to the current segment.
4. Iterate steps 2 and 3 until no more points can be added to the segment.
5. Remove all points of the current segment from the image; repeat steps 2–4 for the next largest local maximum.
6. Repeat steps 2–5 until a user chosen number of segments are found; we find five segments for both positive and negative data.
7. Repeat the process for negative segments by choosing local minima.

It should be noted that as the algorithm iterates, the thresholds adapt to the remaining available data that are not in a previously completed feature. This allows for features with lower maxima than those in features found previously to have a larger spatial extent. In particular, region 2 features, especially the positive current feature, found using adaptive thresholds are more pronounced than when the threshold is fixed. Operationally, the thresholds are chosen as user-defined quantiles of the data.

A3. Choosing Regions 1 and 2 Features

Once a set of spatial segments for each time point has been found, we aim to group and identify these as features in either region 1 or 2 (or neither), depending on the total current in the feature, its area, and its spatial location. This section outlines the procedure by which the features are grouped into region 1 or 2. We give the details for positive currents; the algorithm is applied in the same way to negative currents.

1. Use the image segmentation algorithm to find a given number of positive segments for each time point. We set this number to five; our exploratory results suggest that including more segments does not substantially enhance the final feature.
2. Choose the initial region 1 feature as that with the largest magnitude of the total current among the segments that are more poleward than the average segment detected in step 1.
3. All potential region 2 features are weighted by a weight function that accounts for latitude and angular measures of candidates (details below). The resulting feature with the largest total weighted magnitude is chosen as region 2 feature.
4. A remaining segment is added to region 1 if its area is greater than 1/3 that of the initially chosen feature in step 2 or its current is at least 1/2 that of the initially chosen feature in step 2.

In step 2, the choice of identifying the initial region 1 feature among the higher-latitude segments is to avoid the scenario when the total region 2 current is larger than that of region 1. Although this may sometimes occur, *Coxon et al.* [2014] find that region 1 currents still tend to be larger than those of region 2. In this scenario, this step may fail if too many segments are identified in step 1; we did not run into misidentification problems in the current study, but more study cases are warranted.

The process of choosing region 2 features merits further discussion, as we require information regarding a feature's location and area in addition to current. It is frequently the case that region 1 feature has larger current than region 2 feature, so choosing region 2 features based solely on current often results in miscategorizing region 1 feature. To this end, we introduce a weight function that penalizes unphysical segments. The feature's latitude is given weight as a function of its location relative to the latitude of the data set with the strongest current; features with latitude closer to the highest peaks in the data receive more weight, whereas features farther away from this latitude are given less weight. The angle of the feature is given weight according to the function $(\cos(\theta - \phi) + 1)/2$, where θ is the angle that the centroid of the feature in question makes with the horizontal axis and ϕ is the angle at which the segment falls. This function has a maximum of 1 at ϕ and a minimum of 0 at $\phi - \pi$, so higher weight is given to features with angle measures near ϕ. In our experience, setting $\phi = 0$ works well for most time points when looking for region 2 positive features, and setting $\phi = 3\pi/4$ for the AMPERE data or $\phi = \pi$ for the model output gives good results for region 2 negative features, though the parameter ϕ can be set to a different value for any particular time point. The full set of weights is then found by multiplying both the radial and angular contributions with the area of the corresponding features; each of the positive and negative features with the highest of these weights, not counting region 1 features, is chosen as region 2 features.

Finding region 2 negative currents requires a slight adjustment to the radial component of the weight function, as ringing artifacts in the AMPERE products have the largest impact when choosing region 2

negative feature. To accommodate, we change the radial weighting parameter such that it is higher for smaller latitudes and only consider negative features with latitudes larger than that of region 1 positive feature. We use the area and angle weighting parameters as before.

Two drawbacks to this method of choosing region 2 features are readily apparent. The first is that only one feature is chosen to be in region 2, while there might be another feature in the feature set which is a part of region 2 positive currents; the second is that the angle ϕ is user specified, with the possibility that it may be chosen as time dependent. Developing a method to determine an optimal angle measure for each time point could be a topic of future inquiry.

The algorithm outlined above identifies physically meaningful features in most cases, but there are still times it may fail. We augment the algorithm by temporal smoothing, sharing feature information between adjacent time points. In particular, we introduce two additional weighting parameters to check the features for overlap across past and future time points. These weighting parameters are determined by the number of points in the current time feature that is shared by features in the previous time point and the future time point.

Acknowledgments
This material is based upon work supported by NSF grant DMS-1417724. Part of this research was performed while Sain was at the National Center for Atmospheric Research (NCAR). NCAR is sponsored by the National Science Foundation. We thank the AMPERE team and the AMPERE Science Center for providing the Iridium-derived data products. All data used in this work may be obtained from William Kleiber (william.kleiber@colorado.edu).

References

Anderson, B. J., K. Takahashi, T. Kamei, and C. L. Waters (2002), Birkeland current system key parameters derived from Iridium observations: Method and initial validation results, *J. Geophys. Res.*, *107*(A6), 1079, doi:10.1029/2001JA000080.

Coxon, J. C., S. E. Milan, L. B. N. Clausen, B. J. Anderson, and H. Korth (2014), The magnitudes of the regions 1 and 2 Birkeland currents observed by AMPERE and their role in solar wind-magnetosphere-ionosphere coupling, *J. Geophys. Res. Space Physics*, *119*, 9804–9815, doi:10.1002/2014JA020138.

Heaton, M. J., W. Kleiber, S. R. Sain, and M. Wiltberger (2015), Emulating and calibrating the multiple-fidelity Lyon-Fedder-Mobarry magnetosphere-ionosphere coupled computer model, *J. R. Stat. Soc., Ser. C*, *64*, 93–113.

Iijima, T., and T. A. Potemra (1978), Large-scale characteristics of field-aligned currents associated with substorms, *J. Geophys. Res.*, *83*, 599–615.

Janhunen, P. (1996), GUMICS-3 a global ionosphere-magnetosphere coupling simulation with high ionospheric resolution, in *Environment Modelling for Space-Based Applications*, vol. 392, edited by W. Burke and and T.-D. Guyenne, p. 233, ESTEC, Noordwijk.

Juusola, L., et al. (2014), Statistical comparison of seasonal variations in the GUMICS-4 global MHD model ionosphere and measurements, *Space Weather*, *12*, 582–600, doi:10.1002/2014SW001082.

Kleiber, W., S. Sain, M. Heaton, M. Wiltberger, S. Reese, and D. Bingham (2013), Parameter tuning for a multi-fidelity dynamical model of the magnetosphere, *Ann. Appl. Stat.*, *7*, 1286–1310.

Kleiber, W., S. Sain, and M. Wiltberger (2014), Model calibration via deformation, *SIAM/ASA J. Uncertainty Quant.*, *2*, 545–563.

Korth, H., B. J. Anderson, M. J. Wiltberger, J. G. Lyon, and P. C. Anderson (2004), Intercomparison of ionospheric electrodynamics from the Iridium constellation with global MHD simulations, *J. Geophys. Res.*, *109*, A07307, doi:10.1029/2004JA010428.

Korth, H., B. J. Anderson, J. G. Lyon, and M. Wiltberger (2008), Comparison of Birkeland current observations during two magnetic cloud events with MHD simulations, *Ann. Geophys.*, *26*, 499–516.

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.

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.

Merkin, V. G., B. J. Anderson, J. G. Lyon, H. Korth, M. Wiltberger, and T. Motoba (2013), Global evolution of Birkeland currents on 10 min timescales: MHD simulations and observations, *J. Geophys. Res. Space Physics*, *118*, 4977–4997, doi:10.1002/jgra.50466.

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

Pulkkinen, A., et al. (2013), Community-wide validation of geospace model ground magnetic field perturbation predictions to support model transition to operations, *Space Weather*, *11*, 369–385, doi:10.1002/swe.20056.

Raeder, J., and R. L. McPherron (1998), Global MHD simulations of the substorm current wedge and dipolarization, in *Substorms-4: International Conference on Substorms-4, Lake Hamana, Japan, 9–13 Mar.*, vol. 238, edited by S. Kokubun and and Y. Kamide, p. 343, Kluwer Acad., Dordrecht, London.

Raeder, J., D. Larson, W. Li, E. L. Kepko, and T. Fuller-Rowell (2008), OpenGGCM simulations for the THEMIS mission, *Space Sci. Rev.*, *141*, 535–555.

Rastaetter, 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.

Rastaetter, L., et al. (2013), Geospace environment modeling 2008–2009 challenge: D_{st} index, *Space Weather*, *11*, 187–205, doi:10.1002/swe.20036.

Richmond, A. D., and Y. Kamide (1988), Mapping electrodynamic features of the high-latitude ionosphere from localized observations: Technique, *J. Geophys. Res.*, *93*, 5741–5759.

Richmond, A. D., G. Lu, B. A. Emery, and D. J. Knipp (1998), The AMIE procedure: Prospects for space weather specification and prediction, *Adv. Space Res.*, *22*, 103–112.

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

Tanaka, T. (2000), Field-aligned-current systems in the numerically simulated magnetosphere, in *Magnetospheric Current Systems*, edited by S.-I. Ohtani et al., AGU, Washington, D. C.

Tóth, G., et al. (2012), Adaptive numerical algorithms in space weather modeling, *J. Comput. Phys.*, *231*, 870–903.

Vennerstrom, S., T. Moretto, L. Rastatter, and J. Raeder (2005), Field-aligned currents during northward interplanetary magnetic field: Morphology and causes, *J. Geophys. Res.*, *110*, A06205, doi:10.1029/2004JA010802.

Waters, C. L., B. J. Anderson, and K. Liou (2001), Estimation of global field aligned currents using the Iridium system magnetometer data, *Geophys. Res. Lett.*, *28*, 2165–2168.

Weimer, D. R. (2000), A new technique for the mapping of ionospheric field-aligned currents from satellite magnetometer data, in *Magnetospheric Current Systems*, vol. 118, edited by S.-I. Ohtani et al., pp. 381–388, AGU, Washington, D. C.

Weimer, D. R. (2001), Maps of ionospheric field-aligned currents as a function of the interplanetary magnetic field derived from Dynamics Explorer 2 data, *J. Geophys. Res.*, *106*, 12,889–12,902.

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.

Winglee, R. M., A. T. Y. Lui, R. P. Lin, R. P. Lepping, S. Kokubun, G. Rostoker, and J. C. Samson (1998), Magnetosphere/ionospheric activity during an isolated substorm: A comparison between WIND/Geotail/IMP-8/CANOPUS observations and modeling, in *Geospace Mass and Energy Flow: Results From the International Solar-Terrestrial Physics Program*, vol. 104, edited by J. L. Horwitz and W. K. Peterson, p. 181, AGU, Washington, D. C.