Exploring Predictive Performance: A Reanalysis of the Geospace Model Transition Challenge

D. T. Welling,[1] B. J. Anderson,[2] G. Crowley,[3] A. A. Pulkkinen,[4] L. Rastaetter,[5]

Corresponding author: D. T. Welling, Department of Climate and Space, University of Michigan, 2455 Hayward St., Ann Arbor, Michigan 48109, USA. (dwelling@umich.edu)

[1]Department of Climate and Space,

Abstract. The *Pulkkinen et al.* [2013] study evaluated the ability of five different geospace models to predict surface dB/dt as a function of upstream solar drivers. This was an important step in the assessment of research models for predicting and ultimately preventing the damaging effects of Geomagnetically Induced Currents (GICs). Many questions remain concerning the capabilities of these models. This study presents a reanalysis of the *Pulkkinen et al.* [2013] results in an attempt to better understand the models' per-

University of Michigan, Ann Arbor, Michigan, USA.

[2]The Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland, USA.

[3]Atmospheric and Space Technology Research Associates, Boulder, Colorado, USA

[4]Space Weather Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

[5]Community Coordinated Modeling Center, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

formance. The range of validity of the models is determined by examining the conditions corresponding to the empirical input data. It is found that the empirical conductance models on which global magnetohydrodynamic models rely are frequently used outside the limits of their input data. The prediction error for the models is sorted as a function of solar driving and geomagnetic activity. It is found that all models show a bias towards under-prediction, especially during active times. These results have implications for future research aimed at improving operational forecast models.

1. Introduction

Geomagnetically Induced Currents (GIC) remain an outstanding threat to technological systems on which humans rely. The ultimate source of these currents are the interaction between the solar wind and its embedded interplanetary magnetic field with Earth's magnetosphere and ionosphere [*Boteler*, 2003]. This interaction drives variations in the surface magnetic field, inducing GICs in the conducting lithosphere and any other long conductor, including high-voltage power lines and oil pipelines. The effects of GICs can damage these systems [*Cannon et al.*, 2013] and failures of the power grid, at great economic loss, have been attributed to GICs [*Davidson*, 1940; *Allen et al.*, 1989; *Boteler et al.*, 1998; *Béland and Small*, 2004]. Extreme space weather events could drive GICs with widespread catastrophic consequences on the economy and society [*Committee on the Societal and Economic Impacts of Severe Space Weather Events*, 2008].

A critical step to predicting and eventually mitigating the harmful effects of GICs is predicting the ionospheric currents and variations of the surface magnetic field. Ionospheric and magnetospheric current densities can be combined with accurate knowledge of Earth's impedance to calculate geoelectric field and estimate what portions of the power grid are most at risk [*Pirjola*, 2002; *Thomson et al.*, 2009]. The time rate of change of the surface magnetic field, dB/dt, is a direct indicator of geoelectric field strength [*Viljanen et al.*, 2001]. To this end, it is important to develop, validate, and deploy forecast models into an operational environment. This is one of the major themes of the recently released National Space Weather Action Plan [*National Science and Technology Council*, 2015; *Jonas and McCarron*, 2016].

An important first step towards the realization of operational GIC forecasting was completed by *Pulkkinen et al.* [2013]. This study evaluated five models, both first-principles-based and empirical, that predict surface dB/dt values from upstream solar wind and interplanetary magnetic field (IMF) observations. Data-model comparisons were made between six magnetometer observatories and corresponding model "virtual magnetometers" across six space weather events spanning a range of activity. Binary event analysis [*Jolliffe and Stephenson*, 2003] was used to evaluate each model's ability to forecast when surface dB/dt crossed different activity thresholds. The results of this study were packaged into a report to NOAA's Space Weather Prediction Center that made recommendations concerning which models provided the most operational promise (http://ccmc.gsfc.nasa.gov/challenges/dBdt/).

Despite the advances made by *Pulkkinen et al.* [2013], profound questions about the performance and reliability of these five models persist. Each of these models, at some level, relies on empirical relationships to perform part or all of the calculation. Over what range of space weather conditions are these empirical relationships valid? When do we leave these limits? Furthermore, the metrics used provide single-values to describe performance across all events and activity levels. How does the model error vary with geomagnetic conditions? Do the models perform equally as well during weak, strong, and extreme driving?

This study attempts to answer these questions through reanalysis of the *Pulkkinen et al.* [2013] results. The limits of the empirical inputs are identified for each model, and the distribution of activity for these inputs is compared to the distribution of activity for the six validation events. Data-model error is calculated and binned by solar and geomagnetic

activity to illustrate how each model performs as a function of strength of driving. This analysis is repeated for cases where high and low latitude magnetometers are segregated. These results are compared to illuminate patterns in model behavior and to investigate how the error grows as input conditions exceed the limits of the models' empirical limits.

2. Review of Dataset & Models

To answer the above questions, the results from the *Pulkkinen et al.* [2013] study are evaluated in greater depth. This study tested the ability of five models to predict surface dB/dt over six separate space weather storm events. The events include the famous Halloween Storm of October 29, 2003 [*Skoug et al.*, 2004; *Kappenman*, 2005]. Of the five models tested, two are purely empirical and three are first-principles-based. All data used in this study is freely available at NASA's Community Coordinated Modeling Center (CCMC, http://ccmc.gsfc.nasa.gov/challenges/dBdt). Full details can be found in *Pulkkinen et al.* [2013] and references therein.

The two empirical models include the Weimer Empirical Ground Magnetic Field Prediction Model [*Weimer et al.*, 2010; *Weimer*, 2013] (herein referred to as "Weimer") and the Weigel Empirical Ground Magnetic Field Prediction Model [*Weigel*, 2007; *Pulkkinen et al.*, 2013] (herein referred to as "Weigel"). These models take separate approaches to achieve the same goal: prediction of the perturbation of the surface magnetic field (ΔB) as a function of upstream solar wind drivers. Note that ΔB is the pertubation of the field from background levels, while dB/dt is the time derivative of ΔB. Each uses an extensive input data set to build their empirical relationships, including hundreds of ground-based magnetometer observations. The time range of input data for these models is listed in the top two rows of Table 1.

The three remaining models are different implementations of global magnetohydrodynamic (MHD) models coupled to some ionospheric electrodynamics model. The generic formula for each is to use upstream solar wind observations to drive the model, then perform Biot-Savart integrals of electric currents flowing through the magnetosphere, ionosphere, and "gap-region" (the volume between the inner boundary of the MHD and the height of the ionosphere model) to predict ΔB [*Yu et al.*, 2010]. To obtain horizontal ionospheric currents, a Poison-like equation is solved at a specified altitude using a prescribed conductance distribution. The conductance patterns are some combination of empirical and first-principles models whose inputs are F10.7 solar radio flux, MHD values about the inner boundary, and other parameterizations. The three models are Lyon-Fedder-Mobarry MHD model with the Magnetosphere-Ionosphere Coupler Solver (LFM-MIX, herein referred to as LFM [*Wang et al.*, 2004; *Wiltberger et al.*, 2004]), the Open General Geospace Circulation Model (OpenGGCM, [*Raeder*, 2003; *Raeder et al.*, 2009]), and the Space Weather Modeling Framework (SWMF [*Tóth et al.*, 2005a; *Tóth et al.*, 2012]). While these models have been coupled to different inner magnetosphere models to improve ring current dynamics within the MHD results [*De Zeeuw et al.*, 2004; *Pembroke et al.*, 2012; *Glocer et al.*, 2013; *Yu et al.*, 2014], only the SWMF included this capability in the *Pulkkinen et al.* [2013] study.

The predictions of dB/dt from the five models were compared to observations from six ground-based magnetometers. For the purposes of *Pulkkinen et al.* [2013] and this study, dB/dt refers to the Pythagorean sum of the time derivatives of the two horizontal field components:

$$dB/dt \equiv (dB/dt)_H = \sqrt{(dB_x/dt)^2 + (dB_y/dt)^2} \quad (1)$$

where B_x and B_y are the north-south and east-west components, respectively, of the surface field in geomagnetic dipole coordinates. Of the observatories used, three were high-latitude stations ($> 65°$ geomagnetic latitude) while three were mid-latitude ($> 50°$). Minute resolution data were used, yielding 56,880 data-model pairs per model across all six events. Though the models are capable of producing higher time resolution output (MHD models in particular have a history of investigating faster phenomena [*Claudepierre et al.*, 2009; *Hartinger et al.*, 2014, 2015]), 60 *s* sampling has been shown to capture most of the variation; above these frequencies, the power spectrum falls off considerably [*Pulkkinen et al.*, 2006].

3. Model Range of Validity

The first questions to be addressed are those pertaining to the range of validity of the empirical models tested by *Pulkkinen et al.* [2013]. The range of validity is controlled by model input data: if the model is run using solar wind drivers that do not fall within the range of the solar wind drivers used to create the empirical relationship, there should be no expectation that the model will return an accurate result. This is true because the magnetosphere is a highly non-linear system [e.g., *Price and Prichard*, 1993; *Klimas et al.*, 1996; *Shepherd*, 2007; *Borovsky*, 2014]; it is impossible to know *a priori* how the magnetosphere will behave under increasingly strong driving. By comparing the distribution of solar and geomagnetic activity covered by the six *Pulkkinen et al.* [2013] events to the range of activity covered by empirical input data, judgments about the appropriateness of each model can be made.

The top rows of Figures 1 and 2 illustrate the range of solar wind and geomagnetic conditions covered by six events included in the *Pulkkinen et al.* [2013] study.

These plots show the distribution of minute-resolution solar wind electric field values (VB_Z, in mV/m, Figure 1) and Sym-H index values (nT, Figure 2) across all six of the validation events. These data were obtained from the OMNI database (http://omniweb.gsfc.nasa.gov/ow_min.html). The distributions are described using standard Tukey box plots [*Frigge et al.*, 1989]. The red lines show the median values, the boxes contain 50% of all points and mark the first- and third- quartiles, and the whiskers mark the last point within a distance of 1.5 times the interquartile range of the box. All other points are considered outliers and are indicated by blue dots. The median VB_Z and Sym-H values are 1.28 mV/m and -44 nT, respectively, indicating that the validation events covered predominantly stormy periods. The maximum points of these distributions correspond to the extreme conditions during the Halloween storm of 2003. The width of the whisker diagrams demonstrate that a variety of conditions were covered by the six test events.

The second and third box plots in Figures 1 and 2 show the distribution of VB_Z and Sym-H values for times corresponding to input data for the Weimer and Weigel empirical models (see Table 1). Because the input data time period for both of these models covers the Halloween Storm (Oct. 29, 2003), the VB_Z maxima are the same as those for the SWPC Event distribution (Figure 1, top box plot). The Weigel and Weimer models include the subsequent extreme storms from November 2003, leading to minimum Sym-H values that actually surpass those of the SWPC Sym-H distribution (Figure 2). Because the maxima of the empirical model input data is not surpassed by any point included in the test events, it is known that the *Pulkkinen et al.* [2013] validation study did not exceed the range of validity for these models.

A deeper investigation of these distributions shows that while the data ranges are appropriate, the input data is strongly skewed towards weaker driving and activity compared to the test event conditions. The VB_Z medians for Weimer and Weigel inputs are 0.03 and 0.01 mV/m, respectively, corresponding to almost no solar wind driving. The interquartile ranges of these distributions are miniscule, demonstrating that 50% of all data points used to construct the empirical relationships correspond to periods of negligible driving. The situation is the same for the Sym-H distributions. Comparing to the distributions of activity during the six test events, it becomes clear that the range of conditions tested by the *Pulkkinen et al.* [2013] is not representative of the typical conditions used to create these models. Based on input data alone, it may be expected that these models are more likely to perform well during weak driving, where there is an abundance of input data.

A similar comparison for the three MHD models seems at first unintuitive because these are first-principles, physics-based. However, there are is a key area where the MHD models must turn to empirical relationships: ionospheric conductance. In addition to contributions from solar extreme ultra violet photoionization, which is a slowly varying and well-accounted-for source of conductance, each MHD code attempts to include conductance sources from precipitating electrons. Both the LFM and OpenGGCM models employ a chain of relationships to estimate precipitation from the MHD variables [*Raeder*, 2003; *Wiltberger et al.*, 2009]. These ultimately rely on the *Robinson et al.* [1987] empirical formula to convert estimated electron precipitation to conductance. The SWMF uses a different approach that ties conductance directly to field aligned current strength [*Ridley et al.*, 2001]. This model was built via results from Assimilative Mapping of Ionospheric Electrodynamics (AMIE, *Richmond and Kamide* [1988], which relies on a heavily modi-

fied empirical conductance model [*Spiro et al.*, 1982]). The AMIE results used to build the SWMF empirical conductance model used only magnetometer data as input. Thus, the ionospheric conductance models for each MHD code have a finite range of validity. Conductance is an important value for MHD models [*Ridley et al.*, 2004; *Merkin et al.*, 2005a, b]. It affects both field-aligned and horizontal ionospheric currents, which strongly drive surface ΔB and dB/dt values due to their proximity.

Judging from the range of input data, the empirical conductance models may be an important limitation to the MHD models. The time span of input data for the *Robinson et al.* [1987] relationship is limited to three moderate storm days [*Vickrey et al.*, 1981]. The relationship used in the SWMF was constructed using only a single month of AMIE inputs [*Ridley et al.*, 2001]. The range of solar driving conditions and geomagnetic activity during the time periods for which input data was used are illustrated in Figures 1 and 2, respectively. Note that because the radar observations used to create the *Robinson et al.* [1987] formula were taken in the late 1970s, predating standardized Sym-H, D_{ST} index from the Kyoto World Data Center was used instead. While the median VB_Z values for the MHD empirical conductance models(Figure 1, bottom two box plots) are similar to the medians for the Weimer and Weigel models, the extreme values are far smaller: 2.52 and 2.30 mV/m, corresponding to weak driving. Similarly, the most extreme geomagnetic conditions occurring during the input data time periods are -35 and -65 nT Sym-H/D_{ST}. These values represent weak to moderate driving and correspondingly moderate geomagnetic activity. Clearly, the empirical models leveraged by these MHD models are being exercised well beyond their range of validity.

4. Model Performance vs. Activity

Attention is now shifted to model performance as a function of solar and geomagnetic activity. For every data-model pair, a simple, normalized prediction error is calculated via,

$$Error = \frac{Y - X}{X} \tag{2}$$

where X is the observed value (either ΔB or dB/dt) and Y is the prediction. This normalization allows for the comparison between error during active periods (where perturbations are likely large) and calm periods (where perturbations are likely small). The error values are paired with the corresponding VB_Z and Sym-H values at the time the observation was made. The error is then binned by these values, yielding performance as a function of current space weather conditions. From these distributions, assessments can be made concerning the performance of the various models under different conditions and the connection, if any, to their empirical range of validity.

Figure 3 shows an example result from this analysis applied to the Weimer model. Each box plot represents the distribution of ΔB error values binned by VB_Z. Positive VB_Z values correspond to southward IMF; negative to northward IMF. The box plots are again Tukey box-and-whisker diagrams as in Figures 1 and 2; however, outliers are now excluded. The bin widths are 5 mV/m, ensuring most bins contain > 200 points. Because there are more southward than northward IMF periods covered in the *Pulkkinen et al.* [2013] study, several $VB_Z < 0$ bins contain fewer than 200 points, but all are above 30. A normalized error value of zero is a perfect prediction. Any values above zero are an overprediction: the model predicted a stronger ΔB value than observed. Inversely, any

values below zero are an underprediction, where the model predicted a lower ΔB value than observed.

Figure 3 paints a picture of a model that predicts ΔB reasonably well. Most of the medians (red lines) are slightly below zero, indicating there is a small bias towards underprediction. Despite this bias, the distributions are well clustered about zero with most interquartile ranges (which mark where 50% of the points lie) between ±50% error. Importantly, the spread of the error (measured via the interquartile range and length of the whiskers) is smaller and more regular during southward IMF conditions, when the magnetosphere is most likely to be active. If ΔB error from the Weimer model is instead binned versus Sym-H, the overall pattern does not change much (not shown).

Figure 4 repeats this analysis, but now examines dB/dt error and bins by Sym-H. The bin width is now 25 nT. The distribution of points per bin is more uniform; all bins except the first two contain > 200 points. The comparison to Figure 3 is stark. The medians are consistently well below zero, demonstrating an overwhelming bias towards underprediction. The whiskers do not consistently cross above zero, showing that the likelihood of an overprediction of dB/dt is quite low. This is especially true as Sym-H becomes more negative, indicating stronger geomagnetic activity. A similar pattern emerges if the error is binned by VB_Z (not shown).

Figure 5 shows that the Weimer model is in good company: all of the models included in the *Pulkkinen et al.* [2013] study have a bias towards underpredicting dB/dt (bottom two frames). The median error values, shown now as continuous line plots without the box plot illustrations, are consistently below zero. This pattern emerges regardless of the bin-type, either VB_Z (left hand frames) or Sym-H (right hand frames). For most models,

performance of ΔB prediction (top frames) is better (i.e., closer to zero). However, skill in predicting the instantaneous magnetic perturbation does not appear to translate into accurately capturing time dynamics.

Figure 6 illustrates how these results depend on latitude. The medians of the distributions of normalized ΔB (left column) and dB/dt (right column) are shown when only predictions for high-latitude magnetometers are included (dashed lines) and when only predictions for mid-latitude magnetometers are included (dotted lines). For comparison, the median values for all stations are shown as solid lines (same as in the top row of Figure 5). For the most part, there are not strong differences. However, during strong driving ($VB_Z > 0\ mV/m$) or strong geomagnetic activity (Sym-H$< -200\ nT$), both the OpenGGCM (blue lines) and Weigel (green lines) models show a bias towards overprediction at high latitude stations. Conversely, during the same regions, the SWMF model (orange lines) overpredicts at mid-latitudes while underpredicting at higher locations.

Figure 7 shows a similar comparison, but for normalized dB/dt error. The differences between latitudes are more muted than in Figure 6. Most of the latitude variation comes from the three MHD models, with the two empirical models showing little distinguishing features between high and mid-latitude error results.

5. Discussion

The inescapable conclusion of this study is that all of the models tested in the *Pulkkinen et al.* [2013] have a strong predilection towards underprediction of surface dB/dt. This is best illustrated in the lower right frame of Figure 5: the lines, representing median dB/dt prediction error as a function of geomagnetic activity for each model, only cross zero at three points. This means that an arbitrary prediction from any given model has more

than a 50% chance of yielding a dB/dt value that is lower than observed. As illustrated in Figure 4, the odds can be much, much lower than 50%. Additional model development will be required to produce models that can deliver low-error predictions in an operational environment.

The results of this study reveal why the SWMF (orange lines in Figures 5, 6, and 7) performed most favorably in the work of *Pulkkinen et al.* [2013]. The metrics leveraged in *Pulkkinen et al.* [2013] tested whether a model's prediction crossed a given threshold. Therefore, underpredictions are punished while overpredictions can still count as successful predictions. Of the five models, the SWMF is least likely to underpredict dB/dt (Figure 5). This is especially true during strong solar driving ($VB_Z > 10\ mV/m$) and during periods of strong geomagnetic activity (Sym-H$< -150\ nT$). Such conditions were a focus of the six test events. During periods of northward IMF (lower left-hand frame of Figure 5 where $VB_Z < 0$), the SWMF stands as the only model that is very likely to overpredict dB/dt. These characteristics are desireable when using a binary-event approach towards validation.

An interesting consideration is the impact of the SWMF's inclusion of a ring current model may be having on these results. Inclusion of such a model has been show to strengthen the ring current and region-2 Birkeland currents [*De Zeeuw et al.*, 2004; *Welling et al.*, 2015]. It would be expected that these improvements would affect the mid-latitude predictions most strongly, but still play a role in higher latitude results. Figure 6 appears to support this as the SWMF is the least likely of any model to underpredict ΔB during negative Sym-H conditions (right column). Figure 7 paints a more complicated picture for dB/dt: while the SWMF performs better at high latitudes, the low latitude errors still

closer to zero than any other model. Further investigation is required, but these results suggest that the improved ring current results are reducing prediction error.

This study also illustrates that skill in predicting field perturbation (ΔB) does not directly translate accurate dB/dt. This is most evident with the two empirical models, which perform reasonably well when predicting ΔB, yet severely underpredict the time derivative. This was clearly illustrated for the Weimer model in Figures 3 and 4. The empirical models excel at reproducing the field perturbation, but struggle to accurately reproduce time dynamics. The MHD models, on the other hand, tend to be less likely to underpredict dB/dt while showing no clear advantages in ΔB predictions. It may be possible to compensate for this apparent disconnect by predicting dB/dt directly from the magnitude, and not time derivative, of ΔB, as recently demonstrated by *Tóth et al.* [2014]. It may be necessary to build the empirical models from dB/dt instead of ΔB.

Impulsive changes in the magnetosphere, such as storm sudden commencements, may account for some of the dB/dt performance discrepancy between the MHD and empirical models. SSCs can be clearly captured by the MHD codes. Indeed, during times of positive Sym-H, which are typically the result of compressions of the magnetosphere from impulsive changes in solar wind pressure, the MHD codes separate themselves from their empirical brethren (Figure 5, lower-right frame). The Weimer model averages solar wind inputs over 20 minutes, likely reducing the impact of such impulses. The Weigel model includes an impulsive component, but this was developed to capture dynamics on the order of hours [*Weigel*, 2007]. Because the *Pulkkinen et al.* [2013] study focused on storms, dayside compressions are common and are likely playing a role here.

Other captivating features manifest in these results without readily available explanations. The SWMF, for example, tends to overpredict both ΔB and dB/dt under strongly northward IMF conditions (i.e., negative VB_Z) in grandiose fashion. This feature stems from the commencement and early main phases of the Halloween storm, where the SWMF predicts strong perturbations associated with strongly northward IMF. There is no clear reason why this pattern occurs. Additionally, there is a tendency for all models to underpredict dB/dt during moderate conditions compared to quiet and extreme conditions. Again, there is no obvious explanation for this behavior, and further research into the models' behavior is required.

Candidate explanations for this behavior may be found by exploring limitations of this study and that of *Pulkkinen et al.* [2013]. Uncertainty in the propagation of solar wind drivers from L1 to the nose of the magnetosphere may be playing a role in the timing and amplitude of dB/dt peaks. Studies have shown that the majority of propagation errors are within ± 5 minutes [*Case and Wild*, 2012; *Hassan et al.*, 2015]. Reconnection rate in ideal MHD models is always a point of concern. This is especially applicable in the tail, where explosive reconnection is a requirement for capturing realistic substorm dynamics which contribute strongly to nightside dB/dt. This is less of a concern on the dayside, as MHD models have shown the ability to successfully mimic Petscheck reconnection rates [*Ouellette et al.*, 2010, 2013] and reproduce the immediate ionospheric consequences of dayside reconnection: cross polar cap potential (CPCP) dynamics and region 1 FAC patterns [*Ridley et al.*, 2002; *Korth et al.*, 2008, 2011]. The MHD models have many settings and parameters that affect the results [e.g., *Ridley et al.*, 2010], including grid resolution. The MHD models leveraged by *Pulkkinen et al.* [2013] were configured to run

faster than real time using a modest computational allocation; it is unclear how the results here would be affected if the models used different configurations or higher resolutions. All of these limitations require further research to fully understand.

Finally, these results are placed back into the context of the range of validity. For the two empirical models, there is little evidence that the input data bias towards weak driving and weak activity is impacting model performance. The Weimer and Weigel model error does not grow significantly or systematically as a function of activity. The picture grows even murkier when turning to the empirical conductance models embedded within the MHD models. If the range of validity of the conductance models were indicated on the various error plots, they would cover only a few of the VB_Z or Sym-H bins (incidentally, it is noteworthy that the error medians of the three MHD models tend to converge around the x-axis origins). Additionally, many assumptions and parameters feed into the calculation of precipitation and conductance within MHD models. The full description of the approach leveraged by the LFM model includes four different parameters chosen to optimize a limited number of simulations [*Wiltberger et al.*, 2009]. This makes drawing concrete conclusions about the *Robinson et al.* [1987] model very difficult as it is only one step in the entire process. Despite these confounding factors, expanding the range of validity for the conductance models by including data from geomagnetically active periods will almost certainly have a positive effect on the error patterns. Indeed, studies that change only the conductance model show strong differences in cross polar cap potentials [*Merkin et al.*, 2007]. Updating conductance models should be considered a priority by the community. Alternatively, MHD coupling to advanced, physics-based ionosphere-thermosphere models (a capability that exists in different forms [*Raeder*, 2003; *Ridley*

et al., 2003; *Tóth et al.*, 2005b; *Merkin and Lyon*, 2010]), should be developed, leveraged, and validated.

6. Conclusions

This study performed a reanalysis of the *Pulkkinen et al.* [2013] validation effort. It determined the range of validity for the five models evaluated and compared them to the range of activity used to exercise the codes. Model error for ΔB and dB/dt was binned by solar driving (VB_Z) and geomagnetic activity (Sym-H) to explore how this error changes as a function of those values. This analysis was repeated for high- and mid-latitude magnetometers in isolation. From this analysis, the following conclusions were made:

- Because their input data is drawn primarily from periods of weak activity, the conductance models upon which the MHD codes rely are frequently being used outside their range of validity.
- All models included in the *Pulkkinen et al.* [2013] show a bias towards underprediction of dB/dt. This behavior appears unrelated to performance of ΔB prediction.
- The SWMF is the least likely to underpredict dB/dt, which explains this model's superior performance in the *Pulkkinen et al.* [2013] study. The inclusion of a ring current model appears to be a contributing factor.

These results have implications for future research aimed at improving operational forecast models. The range of validity analysis paints the MHD empirical conductance models as liabilities. Investigating the impact of improvements or replacements will provide great insight into MHD capabilities. Similarly, a deeper study of MHD coupling to ring current models will reveal how important this feature is to both high and mid-latitude predictions.

Determining the relationship between the more accurate ΔB predictions and dB/dt, especially concerning the empirical models, may yield fast improvements to dB/dt predictions. Preliminary work in this direction has shown promise [*Tóth et al.*, 2014]. Finally, it is important to remember that these results are based on six event studies. Validation is an ongoing process and expanding the number of data-model comparisons is critical for a complete understanding of performance and limitations.

Acknowledgments. This project is a result of the efforts of the Geomagnetically Induced Currents Working Group, a NASA Living With a Star Institute. Solar wind and geomagnetic index data was provided by the OMNI database at http://omniweb.gsfc.nasa.gov. Model and observations used in this and previous studies can be obtained from the NASA Community Coordinated Modeling Center at http://ccmc.gsfc.nasa.gov/challenges/dBdt. The results presented in this paper rely on the data collected at various geomagnetic observatories. We thank these institutions for supporting its operation and INTERMAGNET for promoting high standards of magnetic observatory practice (www.intermagnet.org). D_{ST} index was obtained via the Kyoto World Data Center (http://wdc.kugi.kyoto-u.ac.jp)

References

Allen, J., H. Sauer, L. Frank, and P. Reiff (1989), Effects of the March 1989 solar activity, *Eos, Transactions American Geophysical Union*, *70*(46), 1479, doi:10.1029/89EO00409.

Béland, J., and K. Small (2004), Space Weather Effects on Power Transmission Systems: The Cases of Hydro-Québec and Transpower New ZealandLtd, in *Effects of Space Weather on Technology Infrastructure*, pp. 287–299, Kluwer Academic Publishers, Dor-

drecht.

Borovsky, J. E. (2014), Canonical correlation analysis of the combined solar wind and geomagnetic index data sets, *Journal of Geophysical Research: Space Physics*, *119*(7), 5364–5381, doi:10.1002/2013JA019607.

Boteler, D. H. (2003), Geomagnetic Hazards to Conducting Networks, *Natural Hazards*, *28*(2/3), 537–561, doi:10.1023/A:1022902713136.

Boteler, D. H., R. J. Pirjola, and H. Nevanlinna (1998), The effects of geomagnetic disturbances on electrical systems at the Earth's surface, *Advances in Space Research*, *22*(1), 17–27, doi:10.1016/S0273-1177(97)01096-X.

Cannon, P., et al. (2013), Extreme space weather:impacts on engineered systems and infrastructure.

Case, N. A., and J. A. Wild (2012), A statistical comparison of solar wind propagation delays derived from multispacecraft techniques, *Journal of Geophysical Research: Space Physics*, *117*(A2), doi:10.1029/2011JA016946.

Claudepierre, S. G., M. Wiltberger, S. R. Elkington, W. Lotko, and M. K. Hudson (2009), Magnetospheric cavity modes driven by solar wind dynamic pressure fluctuations, *Geophysical Research Letters*, *36*(13), L13,101, doi:10.1029/2009GL039045.

Committee on the Societal and Economic Impacts of Severe Space Weather Events (2008), Workshop Report, *Tech. rep.*, Washington, D.C., doi:10.17226/12507.

Davidson, W. F. (1940), No TitleThe magnetic storm of March 24, 1940Effects in the power system, *Edison Electric Institute Bulletin*.

De Zeeuw, D., S. Sazykin, R. Wolf, T. Gombosi, A. Ridley, and G. Tóth (2004), Coupling of a global MHD code and an inner magnetosphere model: Initial results, *J. Geophys.*

Res., *109*(A12), A12,219, doi:10.1029/2003JA010,366.

Frigge, M., D. C. Hoaglin, and B. Iglewicz (1989), Some Implementations of the Boxplot, *The American Statistician*, *43*(1), 50, doi:10.2307/2685173.

Glocer, A., M. Fok, X. Meng, G. Toth, N. Buzulukova, S. Chen, and K. Lin (2013), CRCM + BATS-R-US two-way coupling, *Journal of Geophysical Research: Space Physics*, *118*(4), 1635–1650, doi:10.1002/jgra.50221.

Hartinger, M. D., D. T. Welling, N. M. Viall, M. B. Moldwin, and A. Ridley (2014), The effect of magnetopause motion on fast mode resonance, *Journal of Geophysical Research: Space Physics*, *119*(10), 8212–8227, doi:10.1002/2014JA020401.

Hartinger, M. D., F. Plaschke, M. O. Archer, D. T. Welling, M. B. Moldwin, and A. Ridley (2015), The global structure and time evolution of dayside magnetopause surface eigenmodes, *Geophysical Research Letters*, *42*(8), 2594–2602, doi:10.1002/2015GL063623.

Hassan, E., S. K. Morley, and J. T. Steinberg (2015), A statistical ensemble for solar wind measurements: A step toward forecasting, in *2015 Los Alamos Space Weather Summer School Research Reports*, edited by M. M. Cowee, pp. 17–31, Los Alamos National Laboratory, Los Alamos.

Jolliffe, I., and D. Stephenson (2003), *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, Wiley, Hoboken, New Jersey.

Jonas, S., and E. D. McCarron (2016), White House Releases National Space Weather Strategy and Action Plan, *Space Weather*, *14*(2), 54–55, doi:10.1002/2015SW001357.

Kappenman, J. G. (2005), An overview of the impulsive geomagnetic field disturbances and power grid impacts associated with the violent Sun-Earth connection events of 29-31 October 2003 and a comparative evaluation with other contemporary storms, *Space*

Weather, *3*(8), doi:10.1029/2004SW000128.

Klimas, A. J., D. Vassiliadis, D. N. Baker, and D. A. Roberts (1996), The organized nonlinear dynamics of the magnetosphere, *Journal of Geophysical Research: Space Physics*, *101*(A6), 13,089–13,113, doi:10.1029/96JA00563.

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, *Annales Geophysicae*, *26*(3), 499–516, doi:10.5194/angeo-26-499-2008.

Korth, H., L. Rastätter, B. J. Anderson, and a. J. Ridley (2011), Comparison of the observed dependence of large-scale Birkeland currents on solar wind parameters with that obtained from global simulations, *Annales Geophysicae*, *29*(10), 1809–1826, doi:10.5194/angeo-29-1809-2011.

Merkin, V., A. Sharma, K. Papadopoulos, G. Milikh, J. Lyon, and C. Goodrich (2005a), Relationship between the ionospheric conductance, field aligned current, and magnetopause geometry: Global MHD simulations, *Planetary and Space Science*, *53*(9), 873–879, doi:10.1016/j.pss.2005.04.001.

Merkin, V. G., and J. G. Lyon (2010), Effects of the low-latitude ionospheric boundary condition on the global magnetosphere, *Journal of Geophysical Research: Space Physics*, *115*(A10), doi:10.1029/2010JA015461.

Merkin, V. G., G. Milikh, K. Papadopoulos, J. Lyon, Y. S. Dimant, A. S. Sharma, C. Goodrich, and M. Wiltberger (2005b), Effect of anomalous electron heating on the transpolar potential in the LFM global MHD model, *Geophysical Research Letters*, *32*(22), doi:10.1029/2005GL023315.

Merkin, V. G., M. J. Owens, H. E. Spence, W. J. Hughes, and J. M. Quinn (2007), Predicting magnetospheric dynamics with a coupled Sun-to-Earth model: Challenges and first results, *Space Weather*, *5*(12), doi:10.1029/2007SW000335.

National Science and Technology Council (2015), National Space Weather Action Plan, *Tech. rep.*

Ouellette, J. E., B. N. Rogers, M. Wiltberger, and J. G. Lyon (2010), Magnetic reconnection at the dayside magnetopause in global Lyon-Fedder-Mobarry simulations, *Journal of Geophysical Research*, *115*(A8), A08,222, doi:10.1029/2009JA014886.

Ouellette, J. E., O. J. Brambles, J. G. Lyon, W. Lotko, and B. N. Rogers (2013), Properties of outflow-driven sawtooth substorms, *Journal of Geophysical Research: Space Physics*, *118*(6), 3223–3232, doi:10.1002/jgra.50309.

Pembroke, A., F. Toffoletto, S. Sazykin, M. Wiltberger, J. Lyon, V. Merkin, and P. Schmitt (2012), Initial results from a dynamic coupled magnetosphere-ionosphere-ring current model, *Journal of Geophysical Research*, *117*(A2), A02,211, doi:10.1029/2011JA016979.

Pirjola, R. (2002), Review On The Calculation Of Surface Electric And Magnetic Fields And Of Geomagnetically Induced Currents In Ground-Based Technological Systems, *Surveys in Geophysics*, *23*(1), 71–90, doi:10.1023/A:1014816009303.

Price, C. P., and D. Prichard (1993), The non-linear response of the magnetosphere: 30 October 1978, *Geophysical Research Letters*, *20*(9), 771–774, doi:10.1029/93GL00844.

Pulkkinen, A., A. Viljanen, and R. Pirjola (2006), Estimation of geomagnetically induced current levels from different input data, *Space Weather*, *4*(8), doi:10.1029/2006SW000229.

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*(6), 369–385, doi:10.1002/swe.20056.

Raeder, J. (2003), Global Magnetohydrodynamics - A Tutorial, in *Space Plasma Simulation*, *Lecture Notes in Physics, Berlin Springer Verlag*, vol. 615, edited by J. Büchner, C. Dum, and M. Scholer, p. 212.

Raeder, J., D. Larson, W. Li, E. L. Kepko, and T. Fuller-Rowell (2009), OpenGGCM Simulations for the THEMIS Mission, in *The THEMIS Mission*, pp. 535–555, Springer New York, New York, NY.

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

Ridley, A., D. L. De Zeeuw, T. I. Gombosi, and K. G. Powell (2001), Using steady-state MHD results to predict the global state of the magnetosphere-ionosphere system, *J. Geophys. Res.*, *106*, 30,067–30,076.

Ridley, A., T. Gombosi, D. L. De Zeeuw, C. Clauer, and A. Richmond (2003), Ionospheric control of the magnetospheric configuration: Neutral winds, *J. Geophys. Res.*, *108*, 2002JA009,464.

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

Ridley, A. J., K. C. Hansen, G. Tóth, D. L. De Zueew, T. I. Gombosi, and K. G. Powell (2002), University of Michigan MHD results of the GGCM metrics challenge, *J. Geophys. Res.*, *107*(A10), 1290, doi:10.1029/2001JA000,253.

Ridley, A. J., T. I. Gombosi, I. V. Sokolov, G. Tóth, and D. T. Welling (2010), Numerical considerations in simulating the global magnetosphere, *Annales Geophysicae*, *28*(8), 1589–1614, doi:10.5194/angeo-28-1589-2010.

Robinson, R., R. Vondrak, K. Miller, T. Dabbs, and D. Hardy (1987), On calculating ionospheric conductances from the flux and energy of precipitating electrons, *J. Geophys. Res.*, *92*, 2565.

Shepherd, S. G. (2007), Polar cap potential saturation: Observations, theory, and modeling, *Journal of Atmospheric and Solar-Terrestrial Physics*, *69*(3), 234–248, doi:10.1016/j.jastp.2006.07.022.

Skoug, R. M., J. T. Gosling, J. T. Steinberg, D. J. McComas, C. W. Smith, N. F. Ness, Q. Hu, and L. F. Burlaga (2004), Extremely high speed solar wind: 2930 October 2003, *Journal of Geophysical Research*, *109*(A9), A09,102, doi:10.1029/2004JA010494.

Spiro, R. W., P. H. Reiff, and L. J. Maher (1982), Precipitating electron energy flux and auroral zone conductances-An empirical model, *Journal of Geophysical Research*, *87*(A10), 8215, doi:10.1029/JA087iA10p08215.

Thomson, A., A. McKay, and A. Viljanen (2009), A review of progress in modelling of induced geoelectric and geomagnetic fields with special regard to induced currents, *Acta Geophysica*, *57*(1), 209–219.

Tóth, G., et al. (2005a), Space weather modeling framework: A new tool for the space science community, *J. Geophys. Res.*, *110*, doi:doi:10.1029/2005JA011126.

Tóth, G., et al. (2005b), Space weather modeling framework: A new tool for the space science community, *J. Geophys. Res.*, *110*, A12,226, doi:10.1029/2005JA011126.

Tóth, G., et al. (2012), Adaptive numerical algorithms in space weather modeling, *Journal of Computational Physics*, *231*(3), 870–903, doi:10.1016/j.jcp.2011.02.006.

Tóth, G., X. Meng, T. I. Gombosi, and L. Rastätter (2014), Predicting the time derivative of local magnetic perturbations, *Journal of Geophysical Research: Space Physics*, *119*(1), 310–321, doi:10.1002/2013JA019456.

Vickrey, J., R. Vondrak, and S. Matthews (1981), The diurnal and latitudinal variation of auroral zone ionospheric conductivity, *J. Geophys. Res.*, *86*, 67.

Viljanen, A., H. Nevanlinna, K. Pajunpää, and A. Pulkkinen (2001), Time derivative of the horizontal geomagnetic field as an activity indicator, *Annales Geophysicae*, *19*(9), 1107–1118.

Wang, W., M. Wiltberger, A. Burns, S. Solomon, T. Killeen, N. Maruyama, and J. Lyon (2004), Initial results from the coupled magnetosphere – ionosphere – thermosphere model: thermosphere – ionospheric responses, *J. Atmos. Sol-Terr. Phys.*, *66*, 1425.

Weigel, R. S. (2007), Solar wind time history contribution to the day-of-year variation in geomagnetic activity, *Journal of Geophysical Research: Space Physics*, *112*(A10), doi:10.1029/2007JA012324.

Weimer, D. R. (2013), An empirical model of ground-level geomagnetic perturbations, *Space Weather*, *11*(3), 107–120, doi:10.1002/swe.20030.

Weimer, D. R., C. R. Clauer, M. J. Engebretson, T. L. Hansen, H. Gleisner, I. Mann, and K. Yumoto (2010), Statistical maps of geomagnetic perturbations as a function of the interplanetary magnetic field, *Journal of Geophysical Research: Space Physics*, *115*(A10), doi:10.1029/2010JA015540.

Welling, D. T., V. K. Jordanova, A. Glocer, G. Toth, M. W. Liemohn, and D. R. Weimer (2015), The two-way relationship between ionospheric outflow and the ring current, *Journal of Geophysical Research: Space Physics*, *120*(6), 4338–4353, doi: 10.1002/2015JA021231.

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

Wiltberger, M., R. S. Weigel, W. Lotko, and J. a. Fedder (2009), Modeling seasonal variations of auroral particle precipitation in a global-scale magnetosphere-ionosphere simulation, *Journal of Geophysical Research*, *114*(A1), A01,204, doi:10.1029/2008JA013108.

Yu, Y., A. J. Ridley, D. T. Welling, and G. Tóth (2010), Including gap region field-aligned currents and magnetospheric currents in the MHD calculation of ground-based magnetic field perturbations, *Journal of Geophysical Research (Space Physics)*, *115*(A14), A08,207, doi:10.1029/2009JA014869.

Yu, Y., V. Jordanova, D. T. Welling, B. Larsen, S. G. Claudepierre, and C. Kletzing (2014), The role of ring current particle injections: Global simulations and Van Allen Probes observations during 17 March 2013 storm, *Geophysical Research Letters*, *41*(4), 1126–1132, doi:10.1002/2014GL059322.

Model Name	Input Observations	Date Range
Weigel	Ground-based magnetometers	Jan. 2000 - Dec. 2006 (7 years)
Weimer	Ground-based magnetometers	Feb. 1998 - Dec. 2005 (8 years)
SWMF	AMIE-Ridley	Jan. 1997 (1 month)
LFM & OpenGGCM	Chatanika radar	Nov. 13, 17 1976 & April 6, 1977 (3 days)

Table 1. Time coverage (right most column) of input data (center column) for the various empirical models (leftmost column) ranges from comprehensive (Weigel and Weimer models, top two rows) to very sparse (SWMF, LFM, and OpenGGCM models, bottom two rows).

Figure 1. Distributions of solar wind electric field (VB_Z, mV/m) for every minute during the six validation events (top box plot, labeled "SWPC") and for every minute corresponding to empirical input data for the Weimer & Weigel models (2nd and 3rd box plots) and for the empirical conductance models used by the MHD models (bottom two box plots). Box plots show median values (red lines), first- and third-quartiles (boxes), and range of data excluding outliers (whiskers). Outliers, defined as any point outside of the box by a distance of 1.5 times the interquartile range, are shown as dots. Positive VB_Z values indicate times of southward IMF, negative values indicate northward IMF. Minimum, maximum, and median values are listed in grey.

Figure 2. Same as Figure 1 but for Sym-H index values. Data is arranged such that more negative values, which indicate periods of stronger geomagnetic activity, are to the right. For time periods that predate the standard Sym-H index, Kyoto D_{ST} is used instead.

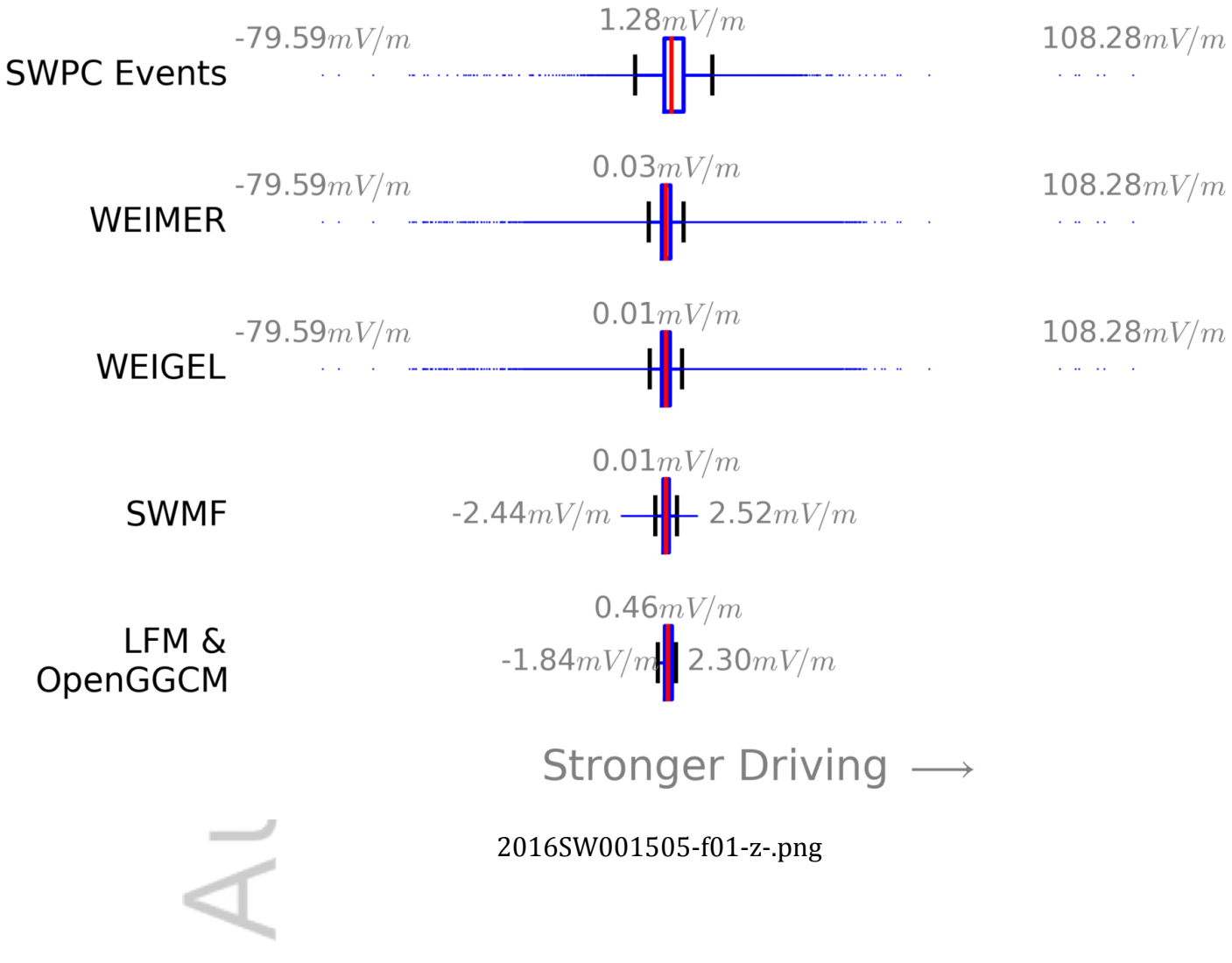
Figure 3. Normalized ΔB error distributions (y-axis and box plots) as a function of solar wind electric field (x-axis, mV/m) for the Weimer model. Box plot interpretation follows that of Figures 1 and 2. Positive error values represent overpredictions, negative values represent underpredictions.

Figure 4. Similar to Figure 3 but for normalized dB/dt error as a function of Sym-H. The x-axis is arranged so that negative Sym-H values, which correspond to higher geomagnetic activity, are to the right.

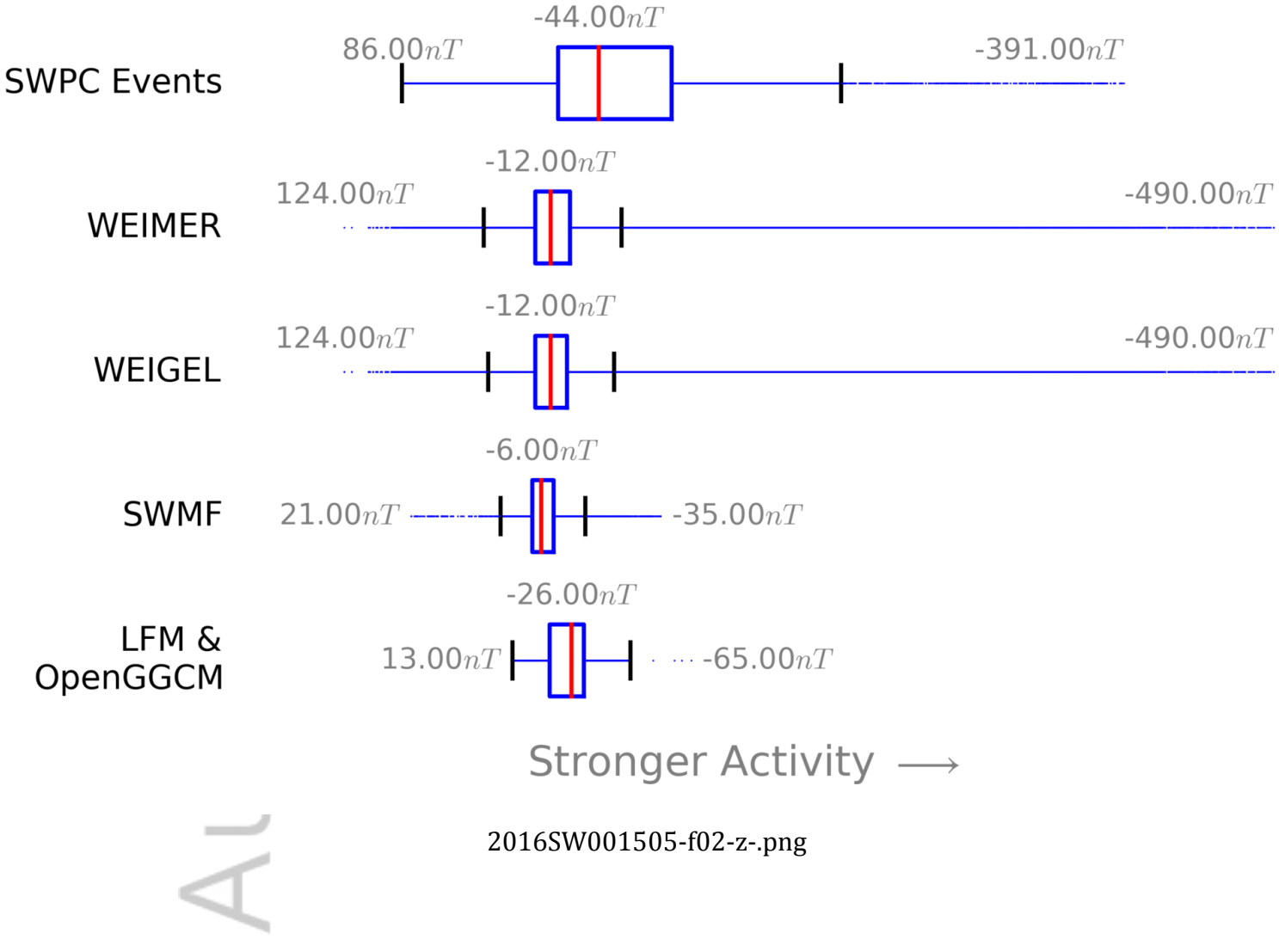
Figure 5. Comparison of median error values, both ΔB (top row) and dB/dt (bottom row), for all models, as a function of VB_Z (left hand frames) and Sym-H (right hand frames).

Figure 6. Comparison of median ΔB errors at high latitude stations (dashed lines), mid latitude stations (dotted lines) and all stations combined (solid lines) for all models (different colors/frames). Left hand column shows median error as a function of VB_Z, right hand column shows error as a function of Sym-H.

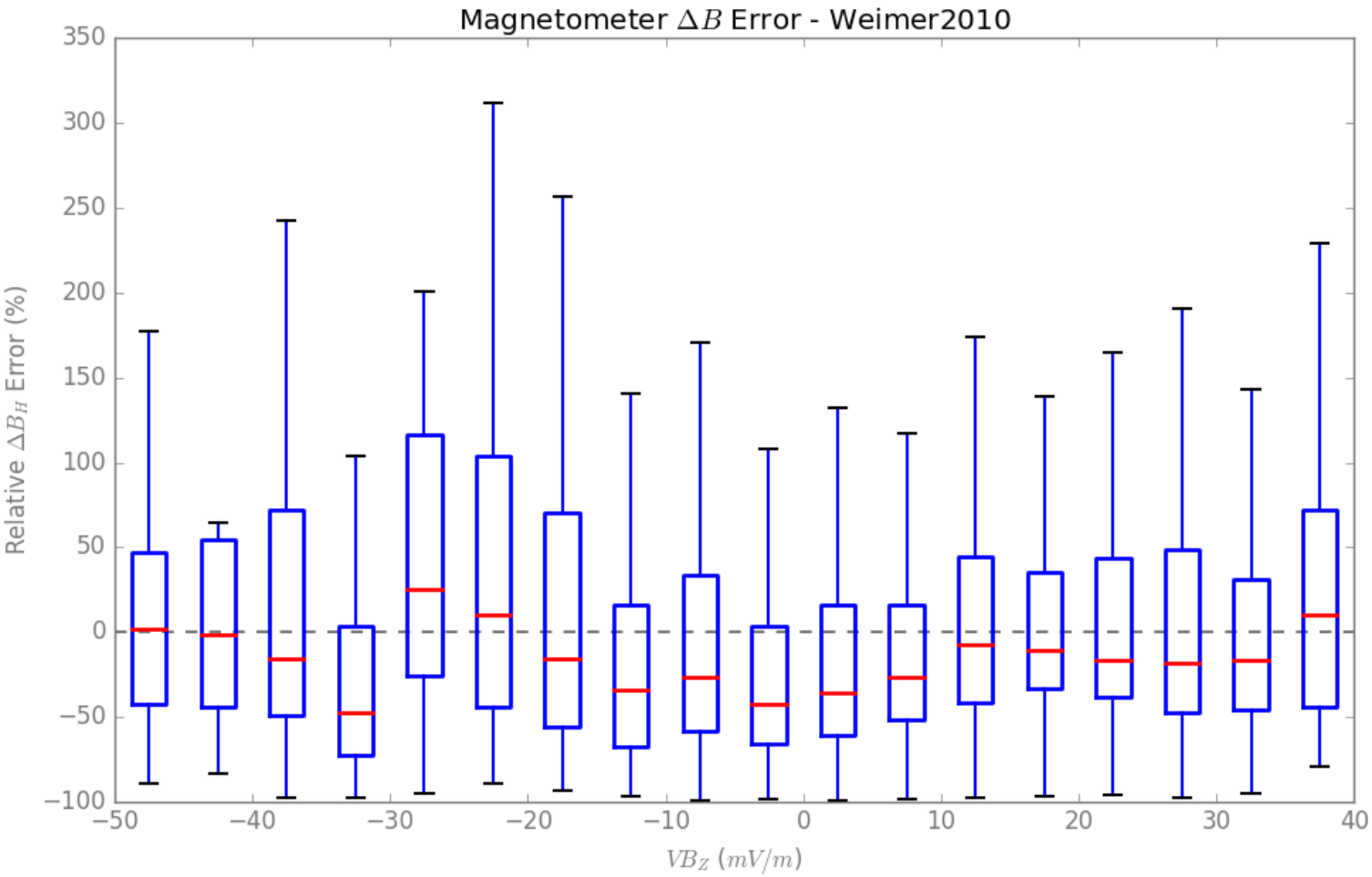
Figure 7. Same as Figure 6 but for dB/dt error.



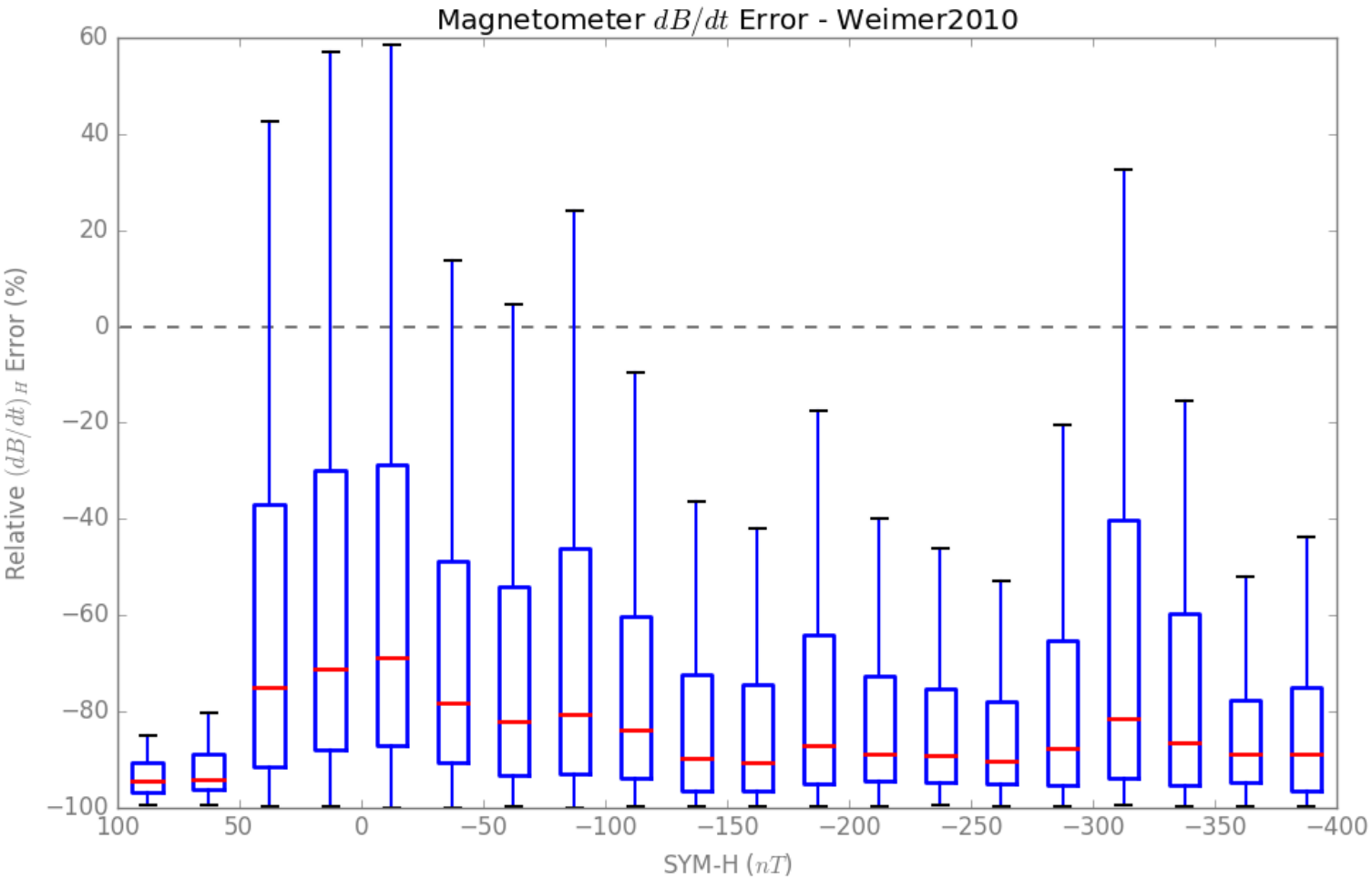
2016SW001505-f01-z-.png



2016SW001505-f02-z-.png



2016SW001505-f03-z-.png



2016SW001505-f04-z-.png