AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Space Physics

RESEARCH ARTICLE

10.1002/2013JA019456

Key Points:
- Predicting local magnetic perturbation dB is easier than predicting dB/dt
- The local dB/dt is highly correlated with dB itself
- The dB/dt can be better predicted indirectly from dB and the correlation

Correspondence to:
G. Tóth,
gtoth@umich.edu

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

Received 18 SEP 2013
Accepted 7 JAN 2014
Accepted article online 10 JAN 2014
Published online 29 JAN 2014

Predicting the time derivative of local magnetic perturbations

Gábor Tóth[1], Xing Meng[1], Tamas I. Gombosi[1], and Lutz Rastätter[2]

[1]Department of Atmospheric Oceanic and Space Science, University of Michigan, Ann Arbor, Michigan, USA, [2]Space Weather Laboratory, Code 674, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

Abstract Some of the potentially most destructive effects of severe space weather storms are caused by the geomagnetically induced currents. Geomagnetically induced currents (GICs) can cause failures of electric transformers and result in widespread blackouts. GICs are induced by the time variability of the magnetic field and are closely related to the time derivative of the local magnetic field perturbation. Predicting dB/dt is rather challenging, since the local magnetic perturbations and their time derivatives are both highly fluctuating quantities, especially during geomagnetic storms. The currently available first principles-based and empirical models cannot predict the detailed minute-scale or even faster time variation of the local magnetic field. On the other hand, Pulkkinen et al. (2013) demonstrated recently that several models can predict with positive skill scores whether the horizontal component of dB/dt at a given magnetometer station will exceed some threshold value in a 20 min time interval. In this paper we investigate if one can improve the efficiency of the prediction further. We find that the Space Weather Modeling Framework, the best performing among the five models compared by Pulkkinen et al. (2013), shows significantly better skill scores in predicting the magnetic perturbation than predicting its time derivative, especially for large deviations. We also find that there is a strong correlation between the magnitude of dB/dt and the magnitude of the horizontal magnetic perturbation itself. Combining these two results one can devise an algorithm that gives better skill scores for predicting dB/dt exceeding various thresholds in 20 min time intervals than the direct approach.

1. Introduction

Geomagnetically induced currents (GICs) were recognized as a major hazard for high-voltage electric systems, such as power lines and transformers [*Boteler et al.*, 1998; *Pirjola*, 2005; *North American Electric Reliability Corporation GMD Task Force*, 2012]. The *National Research Council* [2008] warned about worst-case scenarios that can result in widespread and long-lasting power outages in North America due to a loss of a large number of high-voltage power transformers. The potential economical losses may reach or even exceed the impacts caused by a major hurricane.

The Space Weather Prediction Center (SWPC) of the National Oceanic and Aeronautic Agency develops and provides forecasting products that can be used to mitigate the various harmful effects of space weather. Recently, SWPC expressed interest in predicting the time derivative of the local magnetic field perturbations dB/dt [*Pulkkinen et al.*, 2013]. The dB/dt quantity is closely related to GICs, which is one of the most important quantities for the end users of SWPC. At the request of SWPC, the Community Coordinated Modeling Center (CCMC) has evaluated the predictive capabilities of several physics-based and empirical models for several geomagnetic events [*Pulkkinen et al.*, 2013]. The evaluation is based on the success of the models predicting if the horizontal component of the time derivative of the local magnetic field $(dB/dt)_H$ exceeds some preset threshold values in a given time period. Several models, including the Space Weather Modeling Framework (SWMF) [*Tóth et al.*, 2005, 2012], achieved consistently positive (better than random) skill scores for the majority of events and magnetometer stations.

In this paper, we investigate if there is a way to further improve the efficiency of the SWMF predictions, which were the best among the models compared in the SWPC study. First, we check how well the SWMF can predict the horizontal component of the magnetic perturbation dB_H, as opposed to its time derivative. Second, we establish that there is a surprisingly strong correlation between the magnitude of dB_H and $(dB/dt)_H$. The correlation coefficient is around 0.8 for 10 min or longer time periods. Finally, we use this correlation to predict $(dB/dt)_H$ indirectly, as opposed to a direct approximation. One may regard this approach as a combination of the first principles-based and empirical modeling. The first principles-based model

predicts the quantity dB_H, and then we use an empirical relationship to predict $(dB/dt)_H$ from it. We also attempt to establish the limitations of the new algorithm.

Section 2 gives a brief summary of the CCMC study, including the investigated events and magnetometer stations. Section 3 describes the numerical models and parameters used from the SWMF. Section 4 demonstrates the new algorithm for a particular storm and station. Section 5 presents the results for all events and stations and compares the new indirect method with the direct approach. We also investigate various options for optimizing the new algorithm. Section 6 analyzes the effects of time discretization on the results. We conclude the paper with a summary of the results and a discussion of further directions.

2. The SWPC Challenge

The SWPC challenge is described in detail in *Pulkkinen et al.* [2013]. To keep this paper more self contained, here we briefly summarize the main points. Five models participated in the challenge, three first principles models, among them the SWMF, and two empirical models. Six geomagnetic events were selected to evaluate the models:

1. 29 October 2003 06:00 UT to 30 October 2003 06:00 UT, $Dst_{\min} = -353$ nT
2. 14 December 2006 12:00 UT to 16 December 2006 00:00 UT, $Dst_{\min} = -139$ nT
3. 31 August 2001 00:00 UT to 1 September 2001 00:00 UT, $Dst_{\min} = -40$ nT
4. 31 August 2005 10:00 UT to 1 September 2005 12:00 UT, $Dst_{\min} = -131$ nT
5. 5 April 2010 00:00 UT to 1 April 2010 00:00 UT, $Dst_{\min} = -73$ nT
6. 5 August 2011 09:00 UT to 6 August 2011 09:00 UT, $Dst_{\min} = -113$ nT

The events were divided into $\Delta t = 20$ min intervals, and for each interval starting at time t the quantity

$$dB/dt_H = \max_{[t,t+\Delta t]} \sqrt{\left(dB_n/dt\right)^2 + \left(dB_e/dt\right)^2} \tag{1}$$

is calculated, where the n and e subscripts denote the north and east components of the magnetic perturbations. The time derivatives are obtained from the 1 min resolution time series of the measured or modeled B_n and B_e values. We also define the horizontal magnetic perturbation

$$dB_H = \max_{[t,t+\Delta t]} \sqrt{dB_n^2 + dB_e^2} \tag{2}$$

where dB denotes the deviation from the quiet time baseline field value for observations [*Pulkkinen et al.*, 2013] and the deviation from the intrinsic dipole field for the simulations. Both quantities are observed and simulated at 12 magnetometer stations. The stations can be grouped into four categories according to their magnetic latitude (shown in parentheses):

1. Very high latitude: HRN (73.9°), IQA (74.0°)
2. High latitude: ABK (66.1°), MEA (61.6°), YKC (68.9°), PBQ/SNK (65.5°, 66.1°)
3. Midlatitude: NEW (54.9°), OTT (55.6°), WNG (54.1°)
4. Low latitude: FRD (48.4°), FRN (43.5°), FUR (48.4°)

Station PBQ was replaced by station SNK in November 2007, so PBQ is used for the first four events, and SNK for the last two events. For the full names and exact locations of the magnetometer stations see *Pulkkinen et al.* [2013, Table 2]. We note that the CCMC study concentrates on three high-latitude (PBQ/SNK, ABK, and YKC) and three midlatitude stations (WNG, NEW, and OTT), but here we will include all 12 stations unless noted otherwise.

The observed and simulated dB/dt_H quantities were both compared against four preset threshold values, namely 0.3, 0.7, 1.1, and 1.5 nT/s. The selected threshold values are representative of the variation of the selected events. If both the observation and the model exceed a given threshold, it is a *hit*, if the observed value is above the threshold but the model does not reach it, it is a *miss*, the opposite case is a *false alarm*; finally, if neither values reach the threshold, it is a *no crossing*. The number of hits, misses, false alarms, and no crossings for a given set of events is denoted by H, M, F, and N, respectively. We also introduce $A = H + M$ as the number of intervals that has observed values *above* the threshold and $B = N + F$ as the number of intervals that remain *below* the threshold.

The following quantities provide useful information about the predictive value of the model. The probability of detection (POD) is defined as

$$\mathrm{POD} = H/A \tag{3}$$

The maximum POD value is 1 when all intervals exceeding the threshold value are correctly predicted, and the minimum is 0 when none of the crossings are predicted. Note that POD is undefined if $A = 0$, i.e., there are no observed values above the threshold at all.

The probability of false detection (POF) is defined as

$$\mathrm{POF} = F/B \tag{4}$$

The maximum POF value is 1 when the model predicts values above threshold for all cases when it did not happen in reality, and POF is 0 when the model does not produce any false alarms. Note that the POF is undefined if $B = 0$, i.e., all time intervals contain observed values above the threshold.

Finally, the Heidke Skill Score provides a single measure of the predictive power of the model:

$$\mathrm{HSS} = \frac{2(HN - MF)}{(H + M)(M + N) + (H + F)(F + N)} \tag{5}$$

The maximum Heidke Skill Score (HSS) value is 1 when the model produces a perfect prediction with no false alarms. The HSS value is positive if the model is more predictive than a random number generator that predicts the right fraction of intervals with values exceeding the threshold. The HSS is negative if the model performs worse than a random prediction. Note that the HSS becomes undefined (0/0) if all or no intervals contain observed values above the threshold, and the model predicts this correctly. This means that the HSS is only meaningful if neither A nor B is zero.

3. Space Weather Modeling Framework Settings

Although the overall idea of combining a first principles-based model with an empirical relationship is general, we use the SWMF as the first principles model throughout this paper. This section describes the model settings for the SWMF as used in the SWPC challenge.

The settings were constrained by the following condition set by SWPC: the model has to run at real time speed on a 64-core cluster using state-of-the-art processors and networking. This limited the highest grid resolution to $1/4\,R_E$ and the number of grid cells to about 1 million in the global magnetosphere (GM) model BATS-R-US [*Powell et al.*, 1999; *Tóth et al.*, 2012]. The GM component was coupled with the inner magnetosphere model RCM [*Wolf et al.*, 1982; *Toffoletto et al.*, 2003] and the ionosphere electrodynamics (IE) model RIM [*Ridley et al.*, 2004]. Having an inner magnetosphere model is important to produce a realistic pressure and corresponding ring current during geomagnetically disturbed times [*De Zeeuw et al.*, 2004].

The inner boundary of the global MHD model is at $2.5\,R_E$, therefore, the computational domain does not include the surface of the Earth where the magnetometer stations are located. The inner magnetosphere model RCM does not solve for the MHD field, in fact, it assumes a pure dipole where the magnetic field information is not provided by BATS-R-US. Therefore, the magnetic perturbations at the magnetometer stations are calculated from the currents using Biot-Savart integrals. There are three contributions taken into account: the currents inside the BATS-R-US domain, the field-aligned currents in the gap region between 1 and $2.5\,R_E$ radial distance, and the Pedersen and Hall currents in the ionosphere electrodynamics model RIM. We note that the contribution of ground currents is neglected. The details of this procedure are discussed in *Yu and Ridley* [2008]. Here we use the magnetic perturbations that were calculated by the scripts developed at CCMC to facilitate direct comparison with *Pulkkinen et al.* [2013]. These scripts calculate the same Biot-Savart integrals as those in the SWMF with some minor differences in the implementation.

4. Predicting dB/dt_H at the ABK Station for Event 6

We choose the high-latitude ABK magnetometer station for the 5 August 2011 storm as a concrete example that demonstrates the main idea of the paper. We show the observed and the SWMF predicted dB/dt_H values in Figure 1. Although the model (Figure 1, green line) shows reasonable agreement with the observations (Figure 1, black line) up to $t = 25.5$ h, the model severely under predicts the dB/dt_H values afterward.

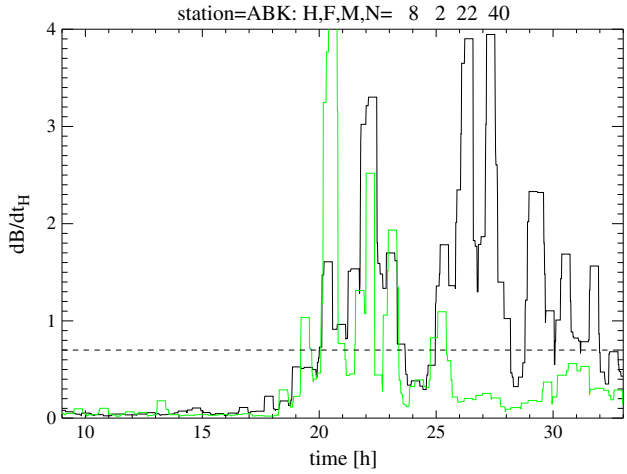

Figure 1. The maximum of the measured (black) and simulated (green) $(dB/dt)_H$ within 20 min intervals at the ABK station for event 6. The dashed line indicates a 0.7 nT/s threshold. H, F, M, and N give the number of hits, false alarms, misses, and no events.

For the 0.7 nT/s threshold (indicated by Figure 1, dashed line) the number of hits, false alarms, misses, and no crossings are $H = 8, F = 2, M = 22$, and $N = 40$, respectively. The large number of misses results in a POD = 0.27 and an HSS = 0.24, while the POF = 0.05 is quite low (few false alarms).

Let us now see how well the SWMF predicts dB_H defined by equation (2). Figure 2 shows the predicted (black line) and observed (green line) values. Comparing Figures 1 and 2, one can see a striking similarity in the overall shapes of the observed curves (black lines). The threshold value of 150 nT (Figure 1, dashed line) was chosen such that it approximately intersects the observed dB_H values about the same time when the observed dB/dt_H crosses the 0.7 nT/s threshold value in Figure 2. Although the agreement between the observed and predicted dB_H is not very good, the predictive power of the model is better than it was for dB/dt_H. The $H = 23, F = 6, M = 8$, and $N = 35$ values show a slight increase in the number of false alarms and a drastic decrease in the number of misses, which results in POD = 0.74, POF = 0.15, and HSS = 0.6, which is much better than it was for the prediction of dB/dt_H.

We continue with examining the correlation between dB/dt_H and dB_H. Figure 3 shows the two quantities plotted against each other on a log-log scale. There is a clear correlation and a roughly linear relationship in the log-log plot. We perform a nonlinear least squares fit using the `lmfit` function of IDL. The fitting function is a power law (corresponding to straight lines on the log-log plot):

$$dB/dt_H = (dB_H/T)^P + E \tag{6}$$

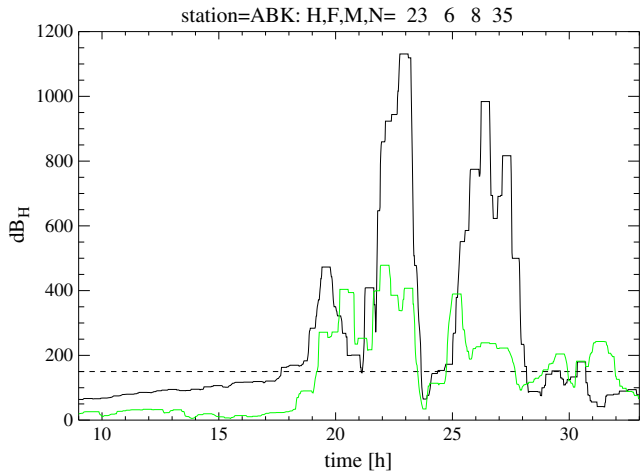

Figure 2. The maximum of the measured (black) and simulated (green) dB_H within 20 min intervals at the ABK station for event 6. The dashed line indicates a 150 nT threshold. H, F, M, and N give the number of hits, false alarms, misses, and no events.

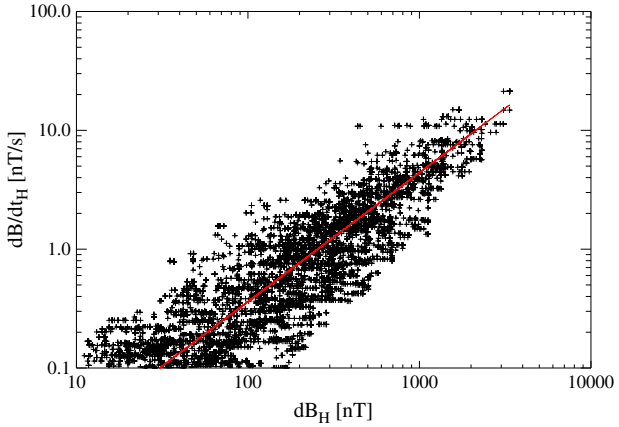

Figure 3. Log-log scatterplot of dB/dt_H versus dB_H as measured by the ABK station. The crosses represent all the 20 min intervals during all the six events. The red line shows the least squares fit $dB/dt_H = (dB_H/255)^{1.08}$.

where T and P are the free parameters, and E is the error that is minimized in the second norm. The best fit is obtained with $T = 255$ and $P = 1.08$. If the power P was exactly 1, we could interpret T as the typical temporal rate of change for dB_H measured in seconds. While it is tempting to make a direct connection between T and the periodicity of the 5 min oscillations [*Kepko and Spence*, 2003], a pure 300 s harmonic oscillation $dB = K \sin(2\pi t/300)$ results in a time derivative $K(2\pi/300)\cos(2\pi t/300)$ so the ratio of maxima over a 20 min interval would be $300/(2\pi) \approx 48$ and not 255. In reality, the magnetic field oscillates with a wide range of periods and amplitudes, and the relationship between dB/dt_H and dB_H results from the combined effect.

Figure 4 shows what happens if we attempt to predict $dB/dt_H > 0.7$ nT/s with the power law function $(dB_H/255)^{1.08}$ of the simulated dB_H. The number of hits $H = 22$ and no crossings $N = 38$ are quite high, while the number of false alarms $F = 4$ and misses $M = 8$ are rather low. The resulting scores POD = 0.73 and HSS = 0.65 are much better than what the model achieved by the direct prediction of dB/dt_H, namely POD = 0.27 and HSS = 0.24. The number of false alarms increased, but POF = 0.1 is still quite low and not much worse than the POF = 0.05 obtained with the direct prediction.

Of course, a single example does not constitute any proof that the indirect algorithm is more successful in general than the direct prediction. The following section explores how the two methods compare for all the stations and all the events.

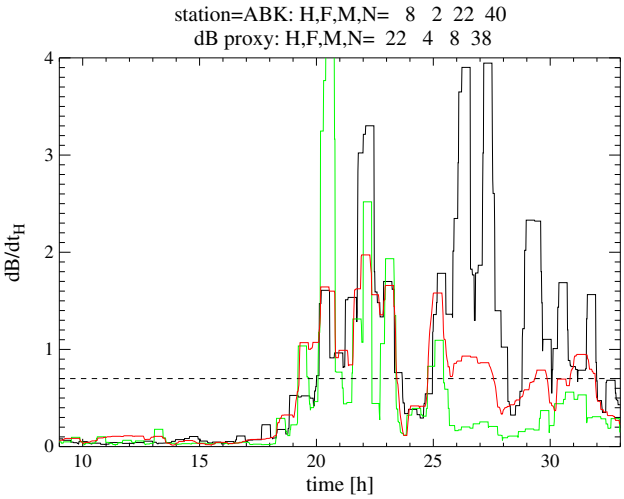

Figure 4. The dB/dt_H observed (black), modeled directly (green), and estimated from the modeled perturbation dB_H as $(dB_H/255)^{1.08}$ (red) for the ABK station and event 6. The dashed line indicates the 0.7 nT/s threshold. The H, F, M, and N values are given for the power law fit (red curve).

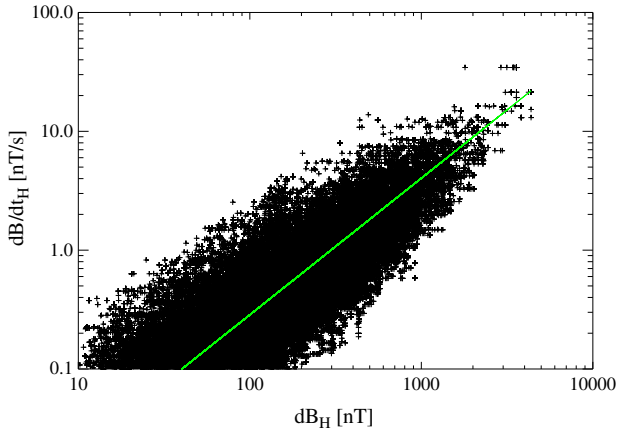

Figure 5. Log-log scatter plot of dB/dt_H versus dB_H measured at all 12 stations. The crosses represent all the 20 min intervals during all the six events. The green line shows the least squares fit $dB/dt_H = (dB_H/301)^{1.16}$.

5. Predicting dB/dt_H Directly Versus Indirectly

Encouraged by the results shown in the previous section, we now proceed and apply the same algorithm for all events and all magnetometer stations. We start with checking how well the power law fit defined by equation (6) works for other stations.

Figure 5 shows the scatterplot for all stations and all events. The power law fit for all stations and all events together gives $T = 302$ and $P = 1.16$, which is not very different from the values found for the ABK station. Table 1 shows the individual least squares fit as well as the correlation coefficients for all stations. The stations are ordered according to their magnetic latitudes. The table shows that the low-latitude stations have the worst correlation coefficients and parameter values that are quite different from the rest. If we exclude the low-latitude stations (FRN, FRD, and FUR), the remaining stations combined give $T = 292$ and $P = 1.14$, i.e.,

$$dB/dt_H = \left(\frac{dB_H}{292\,\mathrm{nT}}\right)^{1.14} \mathrm{nT/s} \tag{7}$$

for the least squares fit. We will use this last fit as the purely data-based empirical relationship between dB_H and dB/dt_H.

The empirical relationship in equation (7) is purely based on measured quantities and uses the data combined for all the stations except for the low-latitude stations. This means that the empirical relationship is not designed to compensate for some systematic deficiencies of the first principles model. For example, if the model would consistently underestimate dB_H, one could reduce the T parameter (and/or increase P) to compensate for it and get a better prediction for dB/dt_H. One could also optimize directly for the skill score instead of using the least squares fit. Finally, one can do these optimizations for each station and

Table 1. Least Squares Fits of $dB/dt_H = (dB_H/T)^P$ for All Events

Station	Magnetic Latitude	T	P	Correlation
FRN	43.5°	819	0.94	0.678
FRD	48.4°	344	1.28	0.706
FUR	48.4°	636	1.00	0.688
WNG	54.1°	334	1.34	0.738
NEW	54.9°	311	1.09	0.840
OTT	55.6°	248	1.04	0.812
MEA	61.6°	229	0.89	0.786
PBQ/SNK	65.5°, 66.1°	277	1.04	0.830
ABK	66.1°	255	1.08	0.842
YKC	68.9°	390	1.21	0.801
HRN	73.9°	277	1.18	0.818
IQA	74.0°	253	1.14	0.858

Table 2. Optimal $dB/dt_H = (dB_H/T)^P$ for Highest HSS Averaged for All Thresholds and Events

Station Group	T	P	HSS Average
Low latitude	327	2.06	0.452
Midlatitude	234	2.49	0.675
High latitude	229	1.21	0.575
Very high latitude	259	1.02	0.474

threshold separately. A typical empirical model does exactly this type of optimizations to provide the best possible prediction.

To establish some "upper limit" on the predictive power of the new algorithm, we optimize the T and P parameters to find the best average Heidke Skill Score (HSS) for various groups of stations. The average is taken over all thresholds and events. Although one could push the optimization even further and optimize for each threshold and station separately, we do not pursue that. We use the `amoeba` procedure of IDL to find the optimum P and T values for each station group. The results are shown in Table 2. There is substantial variation from one station group to another. The T parameters lie in a relatively narrow range from 229 to 327 not very different from the values obtained with the least squares fits to the measurements (Table 1). On the other hand, the exponents are above 2 for the low-latitude and midlatitude station groups, while none of the data-based exponents are above 1.34 in Table 1. This means that the optimization process is now compensating for some of the systematic errors in the model-predicted dB_H.

Figures 6–9 show the event-averaged skill scores as a function of the four different dB/dt_H thresholds for the four different station groups, respectively. The indirect method based on the simulated dB_H and the HSS-optimized power law formula (blue lines) consistently outperforms the direct method based on the simulated dB/dt_H (green lines). Using the same data-based power law fit given by equation (7) also produces superior results (red lines) compared to the direct method except for the smallest threshold $dB/dt = 0.3$ nT/s at the low-latitude and midlatitude stations. For the high and very high latitude stations the red and blue lines are quite close to each other, since the optimized and data-based fits are relatively close to each other.

6. Time Resolution Issues

It is important to investigate how much the time derivative is affected by the 1 min time resolution of the observations and the simulations. Another important question is how much the length of the time intervals, chosen to be $\Delta t = 20$ min so far and used in equations (1) and (2), actually matters. This section examines these problems in some detail.

6.1. Time Resolution of the Data

Since the magnetic field varies quite fast, one may worry if the time derivative is limited by the time resolution of the data. One way to investigate this is to compare two different numerical estimates for the time derivative. The one used in the CCMC study uses the smallest possible stencil:

$$\left.\frac{dB}{dt}\right|_{i+1/2} = \frac{B_{i+1} - B_i}{t_{i+1} - t_i} \tag{8}$$

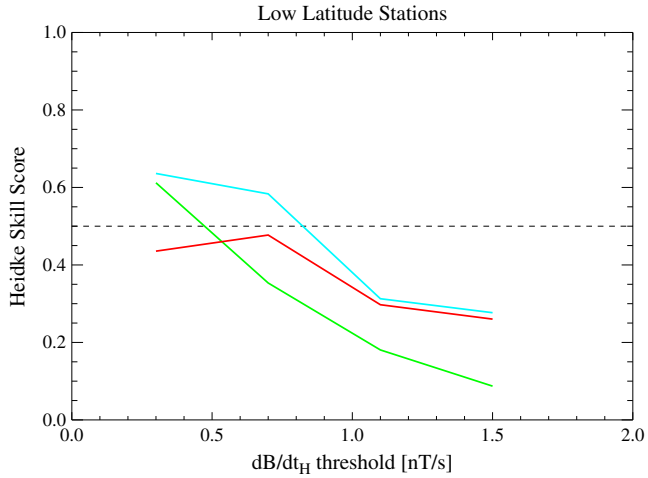

Figure 6. Skill scores averaged over all events as a function of the dB/dt thresholds for the low-latitude stations. The green line is obtained directly from the simulated dB/dt_H. The red line is obtained from the simulated dB_H combined with the data-based $(dB_H/292)^{1.14}$ relationship. The blue line uses the $(dB_H/327)^{2.06}$ formula optimized for the HSS. The dashed line indicates the HSS = 0.5 value for reference.

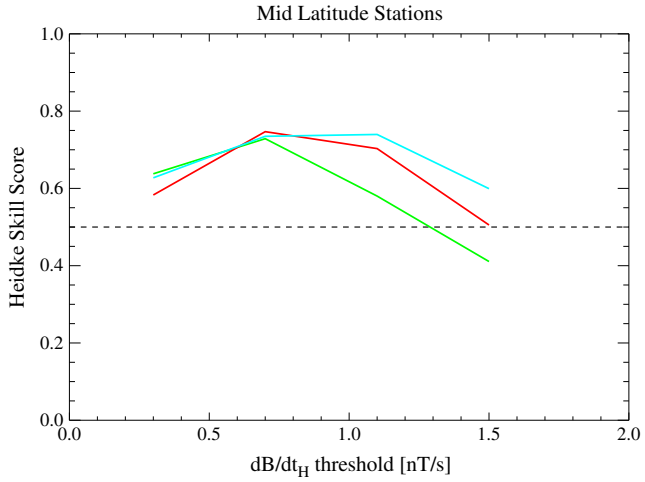

Figure 7. Same as Figure 6 but for the midlatitude stations. The blue line uses the $(dB_H/234)^{2.49}$ fit.

where B is either the north B_n or east B_e component of the magnetic field, and the $i+1/2$ subscript indicates a value taken at time $t_{i+1/2} = (t_{i+1} + t_i)/2$. For the 1 min time series $t_{i+1} - t_i = 60$ s. An alternative formula to calculate the time derivative is

$$\frac{dB}{dt}_i = \frac{B_{i+1} - B_{i-1}}{t_{i+1} - t_{i-1}} \tag{9}$$

which has twice wider stencil. It is easy to see that $dB/dt_i = (dB/dt_{i-1/2} + dB/dt_{i+1/2})/2$, i.e., dB/dt_i can be regarded as a two-cell-wide smoothing of $dB/dt_{i+1/2}$. The horizontal time derivative is obtained in both cases as $dB/dt_H = \sqrt{(dB_n/dt)^2 + (dB_e/dt)^2}$.

Figure 10 (left) compares dB/dt_H obtained with the two different discretizations for station ABK and the first hour of event 1. The difference is quite significant: The difference formula with the narrower stencil produces many higher peaks. This shows that there is significant power in the Fourier spectrum near and below the 2 min periodicity. Since the prediction metric is based on the highest value of the time derivative within an interval, the definition of the time derivative really matters. Figure 10 (right) compares the maximum value of the two curves within a 20 min wide sliding box for the whole event. This approximately represents the function that we are trying to predict. Note that there is a significant difference in the amplitudes. Quantitatively, the average of the 20 min peak amplitudes are 2.96 nT/s and 2.24 nT/s for the narrow and wide stencil discretizations, respectively, so the difference is almost 25%.

Let us repeat a similar analysis for the perturbation dB_H itself. Figure 11 (left) shows the measured $dB_H(t_i)$ in black versus an averaged value $dB(t_{i+1/2}) = [dB_H(t_i) + dB_H(t_{i+1})]/2$ in green at the ABK station for the first hour

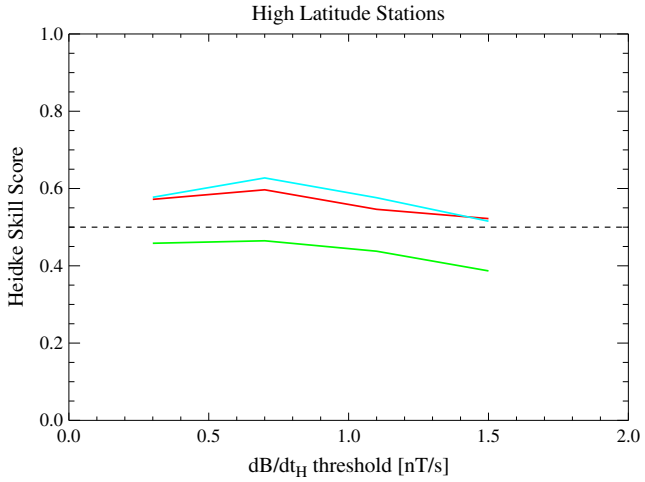

Figure 8. Same as Figure 6 but for the high-latitude stations. The blue line uses the $(dB_H/229)^{1.21}$ fit.

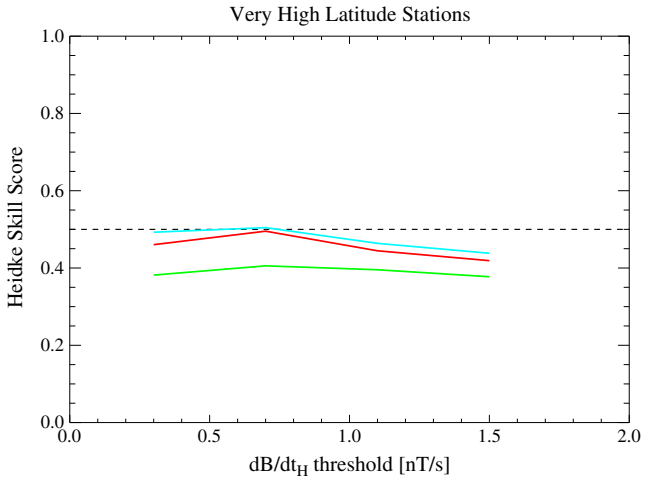

Figure 9. Same as Figure 6 but for the very high latitude stations. The blue line uses the $(dB_H/259)^{1.02}$ fit.

of event 1. The two curves are quite close to each other even at the local maxima. Figure 11 (right) compares the maxima within a 20 min wide sliding interval for the whole event. Again, the 1 min time series and the 2 min average values produce very similar curves. The average of the 20 min peak values are 653 nT and 638 nT, respectively, so the difference is about 2.3% only.

We may conclude that the 1 min time resolution is only marginally sufficient to measure the actual dB/dt_H, and the true maxima are likely to be significantly larger than what is obtained from the finite differences of the 1 min series, even if one uses the narrowest possible stencil. On the other hand, the variation of dB_H itself is reasonably well captured by the 1 min time resolution.

To check the robustness of the results presented in this paper (as well as in *Pulkkinen et al.* [2013]), we redid the calculations using the wide stencil discretization of dB/dt_H. There are some variations in the individual correlation coefficients, power law fits, and skill scores. In particular, the optimal power law fit changes from the one shown in equation (7) to $(dB_H/379)^{1.14}$. The increased scaling constant (379 instead of 292) accounts for the reduced peak amplitudes of dB/dt_H due to the wider discretization stencil, on the other hand the

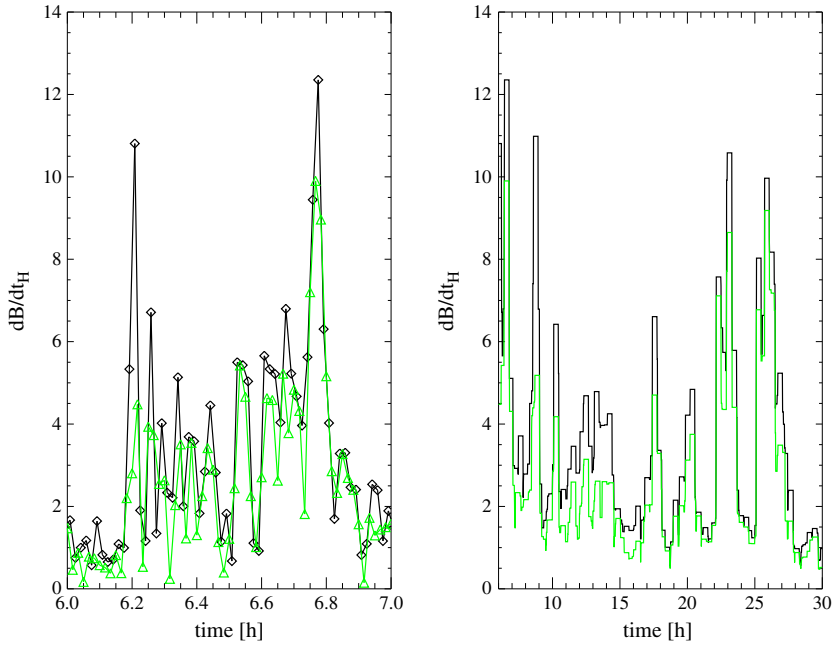

Figure 10. Observed dB/dt_H in (nT/s) at the ABK station for event 1. The black and green curves correspond to the two discretizations in equations (8) and (9), respectively. (left) The 1 min time series for the first hour and (right) the maximum value within a sliding 20 min wide interval for the whole event.

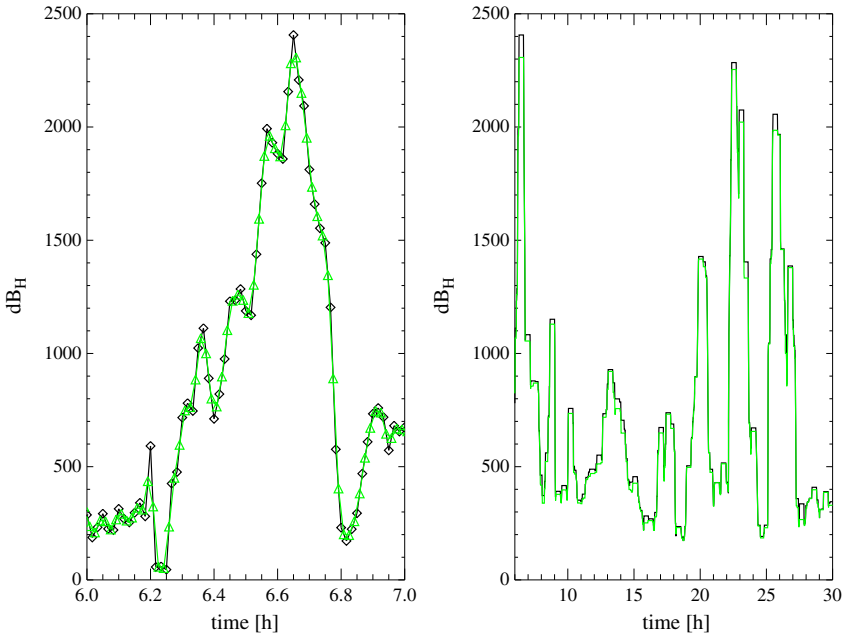

Figure 11. Observed dB_H in (nT) at the ABK station for event 1. The black curve uses the 1 min time series, while the green curve is averaged over two neighboring values. (left) The detailed time series for the first hour and (right) the maximum value within a sliding 20 min wide interval for the whole event.

exponent (1.14) remained the same. Overall, the skill scores do not depend too sensitively whether we use a 1 or 2 min time resolution to obtain dB/dt_H.

6.2. Width of the Time Intervals

The width of the time interval Δt clearly has an effect on predictability. Up to this point we used $\Delta t = 20$ min for the sake of consistency with *Pulkkinen et al.* [2013]. This value is more or less the longest that still makes practical sense, because the simulation based on the kinetically propagated L1 solar wind measurement can only be about 1 h (or less, depending on the solar wind speed as well as the data processing time) ahead of real time. Reducing Δt to smaller values is limited by the time resolution of the predicted magnetic field perturbations (currently set to 1 min), but more importantly, it is limited by the predictive skills of the model. While our model is reasonably successful predicting that dB/dt_H will exceed some threshold within the next 20 min, it is unlikely that it would be successful in predicting the next 5 min or even shorter periods. Neither the direct approach nor the indirect approach (through the correlation with dB_H) are likely to work well for a short time scale, because dB/dt_H is a highly fluctuating quantity as it was shown in the previous subsection.

First, let us examine the correlation between dB/dt_H and dB_H defined in equations (1) and (2) for various values of Δt. We exclude the low-latitude stations as they have the worst correlation and only use the remaining nine stations that were also used to obtain the relationship in equation (7). The time derivative is obtained with the narrow stencil discretization (8), and the magnetic perturbation is interpolated (averaged) to the corresponding time values $t_{i+1/2}$. For the sake of good statistics we use 48 h of observations for all five events, altogether 10 days. The correlation coefficients are 0.86, 0.83, 0.81, 0.79, 0.75, 0.72, and 0.65 for $\Delta t = 40$, 20, 10, 5, 2, 1, and 0 min, respectively. The $\Delta t = 0$ setting means that we use the time

Table 3. HSS as a Function of Δt for Low-Latitude Stations

Threshold	Method	$\Delta t = 1$	2	5	10	20	40
0.3 nT/s	direct	0.274	0.378	0.506	0.562	0.612	0.652
0.3 nT/s	indirect	0.186	0.253	0.334	0.386	0.436	0.490
0.7 nT/s	direct	0.114	0.146	0.198	0.243	0.353	0.465
0.7 nT/s	indirect	0.143	0.201	0.279	0.382	0.477	0.561
1.1 nT/s	direct	0.097	0.135	0.151	0.141	0.181	0.261
1.1 nT/s	indirect	0.066	0.103	0.190	0.246	0.297	0.444
1.5 nT/s	direct	0.024	0.026	0.030	0.071	0.087	0.112
1.5 nT/s	indirect	0.028	0.049	0.133	0.231	0.260	0.329

Table 4. HSS as a Function of Δt for Midlatitude Stations

Threshold	Method	$\Delta t = 1$	2	5	10	20	40
0.3 nT/s	direct	0.479	0.586	0.634	0.651	0.638	0.653
0.3 nT/s	indirect	0.407	0.490	0.533	0.563	0.583	0.613
0.7 nT/s	direct	0.258	0.386	0.559	0.657	0.729	0.735
0.7 nT/s	indirect	0.431	0.551	0.680	0.738	0.747	0.719
1.1 nT/s	direct	0.142	0.199	0.332	0.417	0.580	0.676
1.1 nT/s	indirect	0.251	0.347	0.496	0.587	0.703	0.775
1.5 nT/s	direct	0.079	0.109	0.228	0.304	0.411	0.521
1.5 nT/s	indirect	0.128	0.185	0.279	0.370	0.505	0.643

series dB/dt_H directly and correlate it with dB_H averaged to the same time. Clearly, the correlation is getting worse as Δt is reduced.

Tables 3–6 show how the Heidke Skill Score varies with $\Delta t = 1$ to 40 min for the four dB/dt_H thresholds. The scores are calculated both for the direct prediction of dB/dt_H and for the indirect approach based on equation (7). The four tables correspond to the four groups of stations, sorted by their magnetic latitudes, as shown in Table 1. The results confirm in a quantitative fashion that predicting for shorter time intervals is less successful than for longer time intervals. The dependence of HSS on Δt is monotonic with very few exceptions. It is also clear that the indirect method is better than the direct approach for almost all thresholds and station groups and time intervals. The only exceptions are the low-latitude and midlatitude stations for the smallest threshold 0.3 nT/s, which is the same what was found for the fixed $\Delta t = 20$ min choice. This means that the indirect method is superior (or in a few cases inferior) to the direct method independent of the time interval for the cases examined.

7. Conclusion

The main findings of the paper are the following:

1. The SWMF can predict dB_H better than dB/dt_H.
2. There is a strong correlation between the observed dB_H and dB/dt_H.
3. Using the simulated dB_H and the power law formulas, one can improve the skill scores of predicting dB/dt_H significantly.

The first finding is not very surprising, as it is usually easier to predict a quantity than its time derivative with a first principles-based model. The second finding is also not unexpected as long as there is a typical time scale for the variations of the magnetic perturbations. On the other hand, the strength of the correlation and the goodness of the power law fits are not self evident. The third finding is based on the previous two, but it does not follow. If the dB_H was only slightly better predicted than the dB/dt_H or the correlation was weaker, the new algorithm would not improve the skill scores over the direct approach.

Ultimately, a very good first principles model should produce accurate predictions for dB/dt_H, and the direct method should become optimal. With the currently available models, however, using an empirical power law function of the simulated dB_H provides better skill scores. The improvement is significant enough to be valuable in practice.

Here we did not try to push the combined first principles method and empirical relationship to its limits, although in section 5 we compared the purely data-based empirical relationship (equation (7)) with generic

Table 5. HSS as a Function of Δt for High-Latitude Stations

Threshold	Method	$\Delta t = 1$	2	5	10	20	40
0.3 nT/s	direct	0.352	0.404	0.442	0.469	0.482	0.524
0.3 nT/s	indirect	0.421	0.506	0.564	0.600	0.612	0.629
0.7 nT/s	direct	0.287	0.350	0.428	0.454	0.475	0.508
0.7 nT/s	indirect	0.410	0.499	0.556	0.589	0.606	0.627
1.1 nT/s	direct	0.223	0.288	0.367	0.412	0.451	0.456
1.1 nT/s	indirect	0.341	0.433	0.501	0.539	0.569	0.571
1.5 nT/s	direct	0.181	0.234	0.322	0.376	0.409	0.423
1.5 nT/s	indirect	0.290	0.372	0.457	0.508	0.552	0.570

Table 6. HSS as a Function of Δt for Very High Latitude Stations

Threshold	Method	$\Delta t = 1$	2	5	10	20	40
0.3 nT/s	direct	0.321	0.354	0.381	0.378	0.382	0.401
0.3 nT/s	indirect	0.291	0.362	0.406	0.426	0.460	0.503
0.7 nT/s	direct	0.262	0.313	0.358	0.368	0.404	0.412
0.7 nT/s	indirect	0.274	0.348	0.403	0.441	0.493	0.502
1.1 nT/s	direct	0.222	0.283	0.349	0.386	0.396	0.380
1.1 nT/s	indirect	0.274	0.344	0.399	0.421	0.445	0.452
1.5 nT/s	direct	0.198	0.272	0.333	0.360	0.377	0.374
1.5 nT/s	indirect	0.232	0.318	0.395	0.434	0.419	0.424

power laws (equation (6)) that were optimized for the largest average skill score for various groups of magnetometer stations. As fully expected, the HSS-optimized power law fits give better skill scores than the data-based fit, but the difference is not huge for the majority of stations and thresholds. This suggests that the empirical relationship helps because the first principles model is inherently better in predicting dB_H than dB/dt_H and not because the empirical relationship fixes some systematic error in the model. In fact, we tried to use a power law of the simulated dB/dt_H as a proxy for the predicted dB/dt'_H and optimize for the average HSS. Although there was some slight improvement in the skill scores, it was much less than those shown by the method based on dB_H.

We also found that the 1 min time resolution of the observed data may be insufficient to accurately estimate the true value of dB/dt_H, on the other hand this sampling rate seems to be adequate to estimate dB_H. While dB/dt_H itself changes significantly depending on the time resolution, the overall trends of our results are not too sensitive on how dB/dt_H is calculated. One may be able to calibrate the indirect method to match the dB/dt_H observed with finer time resolution. In practice, one is interested in the geomagnetically induced currents that have a nontrivial dependency on dB/dt_H. Again, combining the simulated dB_H with an empirical relationship between the observed dB_H and GIC amplitudes may be more successful than a direct simulation of GICs.

Finally, we examined the effects of varying the length of the time interval Δt. We found that the prediction skill scores vary monotonically with Δt as expected, i.e., the longer the time interval, the better the prediction can get. We also found that the superiority of the indirect approach over the direct approach is not sensitive to the value of Δt.

Acknowledgments

The simulations used in this paper were performed by Masha Kuznetsova at the CCMC. We also cross validated our scripts against the CCMC scripts for extracting the contingency tables and skill scores from the measured and simulated time series. This proved to be very valuable for both sides, and we thank Antti Pulkkinen for his contributions.

Robert Lysak thanks the reviewers for their assistance in evaluating this paper.

References

Boteler, D. H., R. J. Pirjola, and H. Nevanlinna (1998), The effects of geomagnetic disturbances on electrical systems at the Earth's surface, *Adv. Space Res.*, *22*, 17–27.

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*, A12219, doi:10.1029/2003JA010366.

Kepko, L., and H. E. Spence (2003), Observations of discrete, global magnetospheric oscillations directly driven by solar wind density variations, *J. Geophys. Res.*, *108*(A6), 1257, doi:10.1029/2002JA009676.

National Research Council (2008), *Severe Space Weather Events—Understanding Societal and Economic Impacts: A Workshop Report*, National Academies Press, Washington D. C.

North American Electric Reliability Corporation GMD Task Force (2012), 2012 special reliability assessment interim report: Effects of geomagnetic disturbances on the bulk power system, *Tech. rep.*, NERC.

Pirjola, R. J. (2005), Effects of space weather on high-latitude ground systems, *Adv. Space Res.*, *36*, 2231–2240.

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

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.

Ridley, A., T. Gombosi, and D. Dezeeuw (2004), Ionospheric control of the magnetosphere: Conductance, *Ann. Geophys.*, *22*, 567–584.

Toffoletto, F., S. Sazykin, R. Spiro, and R. Wolf (2003), Inner magnetospheric modeling with the Rice Convection Model, *Space Sci. Rev.*, *107*, 175–196.

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

Tóth, G., et al. (2012), Adaptive numerical algorithms in space weather modeling, *J. Comput. Phys.*, *231*, 870–903, doi:10.1016/j.jcp.2011.02.006.

Wolf, R. A., M. Harel, R. W. Spiro, G. Voigt, P. H. Reiff, and C. K. Chen (1982), Computer simulation of inner magnetospheric dynamics for the magnetic storm of July 29, 1977, *J. Geophys. Res.*, *87*, 5949–5962.

Yu, Y., and A. Ridley (2008), Validation of the Space Weather Modeling Framework using ground-based magnetometers, *Space Weather*, *6*, S05002, doi:10.1029/2007SW000345.