Predictability of Large Geomagnetic Disturbances Based on Solar Wind Conditions

Robert S. Weigel, Daniel N. Baker, E. Joshua Rigler, and Dimitris Vassiliadis

***Abstract*—We test the ability of a data-derived model of geomagnetic activity, originally optimized to have a high prediction efficiency (PE), for its ability to predict only large geomagnetic disturbances. Correlation-based metrics, such as prediction efficiency, are often used as a measure of model performance. This metric puts equal weight on prediction of both large and small measurements. However, for space weather purposes, one is often interested in knowing only if a large disturbance event will occur so less emphasis should be placed on small measurements. If only large events are of interest, then a correlation metric is not the best measure of model performance. In this work, we determine how well a data-derived model, originally optimized to have a high prediction efficiency, predicts large geomagnetic events. The ratio of the number of correct to false alarm forecasts, R_F, is used as an event-predictor metric. It is shown that in the electrojet regions the data-derived model that predicts the north–south component of the ground magnetic field B_x has a spatial R_F profile similar to that of the prediction efficiency. Maximal values of $R_F = 4$ are found at 0300 MLT when an event is defined as an excursion in the hourly-averaged north–south component of the ground magnetic field below -400 nT. Whereas the local time profile of $PE(B_x)$ is similar to $R_F(B_x)$, the profile of $PE(|dB_x/dt|)$ differs substantially from $R_F(|dB_x/dt|)$ in the noon sector. Epoch analysis shows that the poor performance in the noon sector is a result of pre-event levels of $|dB_x/dt|$ not being clearly separated from post-event levels.**

***Index Terms*—Decision-making, Geomagnetism, Geophysical measurements, Geophysical signal processing, Prediction methods.**

I. Introduction

PREDICTION of large geomagnetic disturbance events is often important for real forecasting situations; in many cases, such events require a discrete (or binary) prediction that specifies only if an event will or will not happen. This is in contrast to the fact that most data-derived (or "inverse") geomagnetic models (e.g., linear filters, neural network filters, etc.,) are not optimized to predict events; in most cases, data-derived filters are optimized to have a high data-model correlation or prediction efficiency. That is, the linear or neural network filter parameters are determined such that the model is good at predicting both large and small amplitudes. Supposing that prediction of large events is only of interest, the best approach would be to develop a new data-derived filter with parameters such that the model is optimized to have success in predicting events. Such redevelopment of a filter is not always possible and can be computationally expensive. A more likely scenario is that previously developed models will be used as event predictors.

In this paper, we consider how a previously developed neural network filter model that was originally developed to have a high PE performs in predicting large geomagnetic events. We seek to set a benchmark for the predictability of large geomagnetic events in the same way that [1] set a benchmark for the predictability, in terms of PE, of the amplitude of the ground magnetic field and its time derivative as a function of local time.

The events considered in this paper are crossings of two functions of the ground magnetic field above a threshold value: 1) the hourly average value of the north–south component of the geomagnetic field, $|B_x|$, measured by auroral-zone magnetometer and 2) the hourly average of $|dB_x/dt|$ computed using the one minute differenced B_x data. We first review the ability, in terms of prediction efficiency, of the two neural network prediction filters developed by [1] that predict B_x and $|dB_x/dt|$, respectively. These two models are then evaluated, in terms of the R_F metric, for their ability to predict only large events. The R_F metric is the ratio of the number of correct forecasts to the number of false alarm forecasts. It is a quantity that can be used by a large class of users to determine if always taking mitigating action based on a prediction of an event will have economic utility [2]–[4].

II. Analysis

The data set contains ground magnetometer measurements from the Sodankylä magnetometer station at one-minute resolution for days 23–365 of 1998 and all days of 1999 and 2000. The solar wind magnetic field data (in GSM coordinates) from the MAG instrument on ACE [5] were interpolated from 16-s averages to a 1-min time grid. The solar wind ion-velocity data from the SWEPAM instrument on ACE [6] were interpolated to a 1-min time grid from 64-s averages.

Reference [1] developed two data-derived neural network models and evaluated their ability to predict two measures of geomagnetic fluctuations. One model predicts the 30-min average of the north–south component of the magnetic field B_x, while the other predicts the 30-min average of $|dB_x/dt|$ with $dt = 1$ min. The parameters of the neural network were determined using input variables of six 30-min averages of the solar wind measurements $[v_x, B_y^{\mathrm{IMF}}, B_z^{\mathrm{IMF}}, \rho_i]$ or combinations thereof. Here we review the results of this work focusing only on one of the auroral-zone magnetometer stations, Sodankylä

Manuscript received September 22, 2003; revised January 29, 2004. This material is based upon work supported by CISM, which is funded by the STC Program of the National Science Foundation under Agreement Number ATM-0120950.

R. S. Weigel, D. N. Baker, and E. J. Rigler are with the Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO 80301 USA (e-mail: robert.weigel@lasp.colorado.edu; jrigler@colorado.edu).

D. Vassiliadis is with the Universities Research Corporation at NASA/Goddard Space Flight Center, Seabrook, MD 20706 USA (e-mail: vassi@electra.gsfc.nasa.gov).

Digital Object Identifier 10.1109/TPS.2004.830992

(SOD), which has geographic latitude and longitude of (67.37, 26.63) and CGM latitude and longitude (in 1998) of (63.87, 107.61).

As shown in Fig. 1(a), the PE obtained by the B_x model ranges from zero to 0.7. The PE was also evaluated when only the individual variables of v_x or $B_z^{\rm IMF}$ were the only inputs to the filter. Fig. 1(a) shows that v_x and $B_z^{\rm IMF}$ have similar PE profiles. This indicates that they both have a contribution in predicting variations in B_x. This is expected because one of the main drivers of auroral zone geomagnetic fluctuations is the product of v_x and the southward component of $B_z^{\rm IMF}$.

The locations of highest PE correspond to the average location of the eastward and westward electrojets in the expanded auroral oval. This was interpreted as indicating that the model was capturing solar wind driving of these current systems. In between these regions, the PE is near zero as a result of a combination of factors, including the low average values of the magnetic field amplitude and the lack of a consistent solar wind driven current system appearing in these locations. Fig. 1(b) shows that the PE of $|dB_x/dt|$ is less dependent on spatial location than the PE of B_x. The independent contribution of v_x and $B_z^{\rm IMF}$ has a strong local time dependence, which was interpreted as an indication that the solar wind-$|dB_x/dt|$ coupling function has a MLT dependence.

Having reviewed the ability of the two models to predict all levels of $|B_x|$ and $|dB_x/dt|$ as a function of local time, suppose now that one is only interested in knowing if the disturbance measure $D \equiv B_x$ or $|dB_x/dt|$ will rise above a threshold level D_c. As noted in the introduction, a new model can be developed that is optimized to perform well in predicting only threshold crossing events. Alternatively, a model that predicts all levels of activity can be used for event predictions. In this work, we choose to evaluate the ability of the two previously developed models as event predictors. The output of these two models are transformed into a event prediction by stating that when the filter output, S, predicts a value of D above a given level, a prediction of an event in D is made. We define an event as a crossing in the disturbance signal from below to above D_c and a forecast of an event in D is made at hour t_F when the set of rules

$$S(t_F-1) \geq S_c \,\text{and}\, S(t_F-2) < S_c \,\text{and}\, D(t_F-1) < D_c \quad (1)$$

is satisfied, where S_c is an adjustable threshold parameter. Thus, if the hourly average of S goes from below S_c to above S_c, a prediction of a threshold crossing in D is made for the next hour, provided D was not already above D_c. Given the set of event times t_E from inspection of the D time series and the set of event-forecast times t_F from the S time series, the number of correct forecasts is the number of t_F values for which there is a t_E in the interval $t_F \leq t \leq t_F + 1$. This means that the forecast of an event is extended over 2 h. A similar approach for the prediction of large increases in the daily average of relativistic electron fluxes based on $S = v_x$ has been developed in [8].

The quantity R_F is used as an event-predictor metric. As noted in the introduction, the ratio of the number of correct to false alarm forecasts is an important measure of model performance from a user's perspective. This ratio represents the maximum cost/benefit ratio that a broad class of users can have if taking mitigating action following each forecast is to have positive monetary utility [4]. Other quantities, such as the number of misses, may be important for a different class of user, but we will restrict the analysis to the R_F ratio. The possible outcomes of an event forecast are listed in Table I.

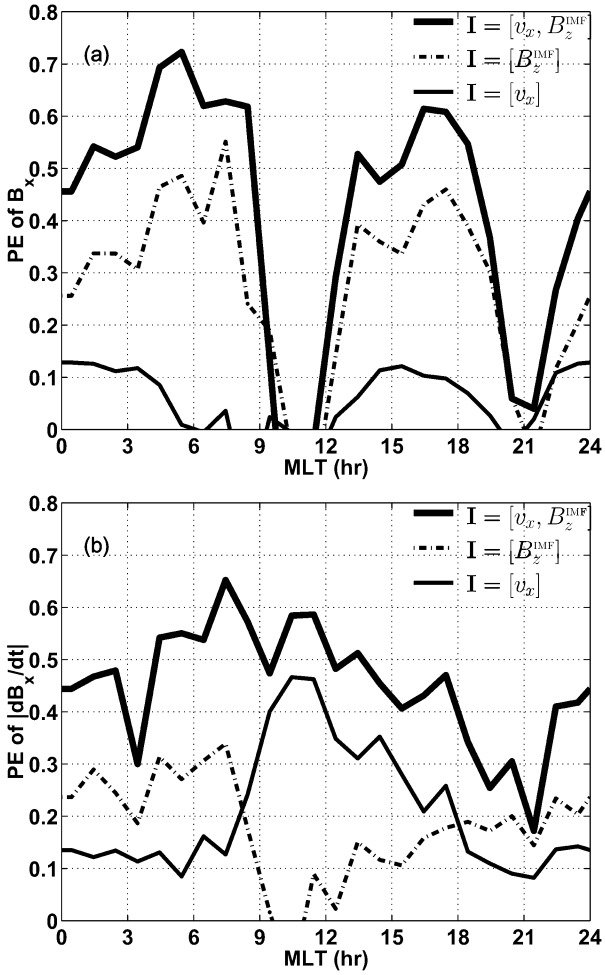

Fig. 1. From [7], the prediction efficiency of models which use different input data $\mathbf{I}$ at auroral-zone station SOD for the PE of (a) B_x and (b) $|dB_x/dt|$. The prediction efficiency was computed by taking all predictions of B_x (or $|dB_x/dt|$) over the three-year interval at a given local time and comparing this with the measurements at that local time.

To analyze the event forecasts at different local times, $R_F \equiv N_H/N_{\overline{H}}$ is computed for all forecasts in a sliding 5-h window centered on a given MLT. This computation was made for S_c values adjusted in increments of 1% of the full range of S. The R_F value reported corresponds to its maximum value with respect to S_c. We have restricted the minimum number of acceptable forecasts to 30; the number of data points we are fitting the model to is the number of forecasts, and this number should be much less than the number of free parameters (one in this case, S_c).

In Fig. 2(a), the ratio of correct to false alarm forecasts is shown for $D = |B_x|$, at three threshold levels, $D_c = 220$, 320, and 400 nT. These values correspond to the 96th, 98th, and 99th percentiles in the local $(\text{MLT} \pm 2\ \text{h})$ probability distribution of $|B_x|$. The profiles are similar to Fig. 1(a), in that the peaks are near the locations of the eastward and westward electrojets. One difference is that the post-midnight sector peak is much higher

TABLE I
CONTINGENCY TABLE FOR AN EVENT FORECAST WITH ELEMENTS OF NUMBER. N_H REPRESENTS THE NUMBER OF CORRECT FORECASTS WHILE $N_{\overline{H}}$ IS THE NUMBER OF FALSE ALARMS. THE NUMBER OF MISSES IS GIVEN BY N_M AND x IS THE NUMBER OF INTERVALS WITH NO FORECASTS OR EVENTS. N_F REPRESENTS THE TOTAL NUMBER OF FORECASTS, WHILE THE TOTAL NUMBER OF TIME INTERVALS WITHOUT A FORECAST IS $N_{\overline{F}}$. NOTE THAT $N_E + N_{\overline{E}}$ AND $N_F + N_{\overline{F}}$ BOTH SUM TO N, THE TOTAL NUMBER OF WARNING PLUS NON-WARNING INTERVALS. THIS TOTAL NUMBER MAY BE LESS THAN THE NUMBER OF POINTS IN THE TIME SERIES BECAUSE FORECASTS ARE EXTENDED OVER TWO DAYS

	Observed		
Forecast	Yes	No	Total
Yes	N_H	$N_{\overline{H}}$	$N_F = N_H + N_{\overline{H}}$
No	N_M	x	$N_{\overline{F}} = N_M + x$
Total	$N_E = N_H + N_M$	$N_{\overline{E}} = N_{\overline{H}} + x$	N

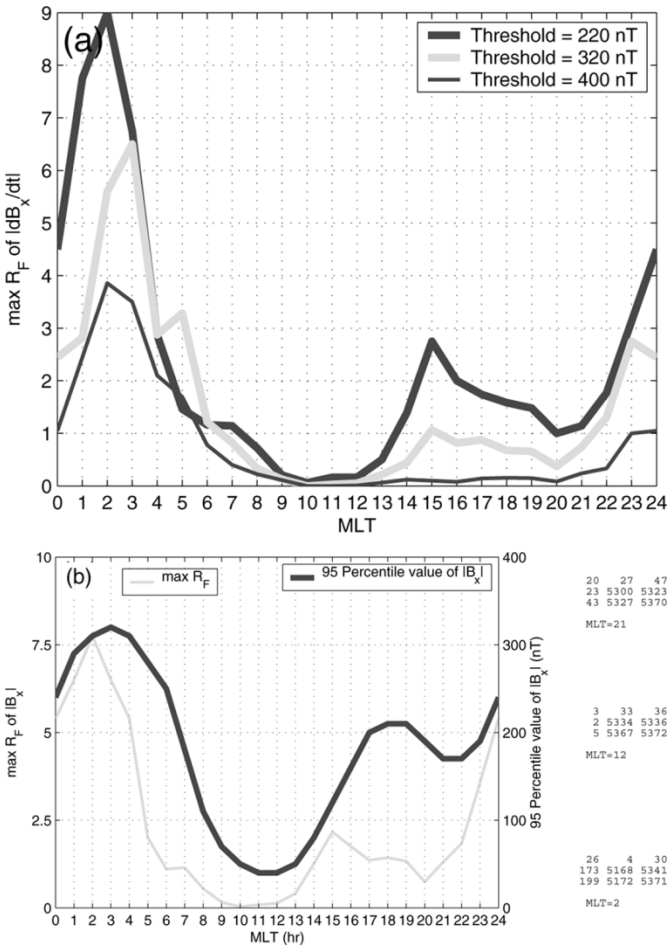

Fig. 2. (a) Ratio of the maximum (with respect to S_c) number of correct to false alarm forecasts of a threshold crossing in $|B_x|$ above threshold values corresponding to the 96, 98, and 99th percentiles of the distribution at that local time. (b) Ratio when the threshold value is chosen to be the 95th percentile value at a given MLT. To the right of the figures, contingency tables in the format of Table I are shown for selected local times.

than the pre-dusk peak. This indicates that for a given threshold value, events in the westward electrojet region are better predicted than those in the eastward electrojet region with respect to the R_F ratio. One contribution to this difference may be a result of the distribution of amplitudes of $|B_x|$ being less extended in the pre-dusk sector. Because of this, there are a fewer number of events for a given threshold in the pre-dusk sector. Therefore, low values of R_F may be due to a relative difference in the number of events where $|B_x|$ exceeds the threshold.

To eliminate the effect of differing number of events at a given local time, we can compute R_F as a function of MLT with a local-time dependent value of D_c chosen so that there are an equal number of events above D_c for each MLT. Although less meaningful from a practical standpoint, because absolute amplitudes are usually of more interest, such analysis allows comparison of the predictability of events that are large with respect to the local distribution. Fig. 2(b) shows the R_F versus MLT profile along with the 95 percentile value of $|B_x|$. Because this R_F curve still has a higher peak in the westward electrojet region, the difference in R_F between the two regions is not due to differing number of events for a given threshold value.

Fig. 2(b) also contains contingency tables for selected MLT values. At $\mathrm{MLT} = 2$, the model was most successful at predicting large excursions in $|B_x|$; of the 30 event forecasts, only 4 were false alarms. At $\mathrm{MLT} = 12$, the model was comparatively unsuccessful at predicting relatively large excursions in $|B_x|$; of the 36 event forecasts, only 3 were correct.

In Fig. 3(a), the ratio of correct to false alarm forecasts is shown for $D = |dB_x/dt|$, at three threshold levels, $D_c = 13\ \mathrm{nT/min}$, 19 nT/min, and 26 nT/min. The curve of $R_F(|dB_x/dt|)$ does not have a profile similar to $PE(|dB_x/dt|)$ in Fig. 1(b). In the noon sector, there is a deep minimum in R_F, while the pre- and post- midnight sectors have high values. Along with other considerations, [1] interpreted Fig. 1(b) as showing that the model was capturing primarily Kelvin-Helmholtz (K-H) driving of $|dB_x/dt|$ in the noon sector and reconnection driving in the midnight sector. With this interpretation and the result shown in Fig. 3(a), we conclude that $|dB_x/dt|$ events of a given amplitude driven by reconnection are more predictable by the model than K-H driven events. Although the model predicted fluctuations driven by these processes an equal amount in terms of PE, the model has a significantly different prediction performance in terms of R_F for events driven by these two processes. In Section III, epoch analysis is used to determine possible reasons for this.

As before, we can test whether the profile of R_F as a function of MLT is influenced by a differing number of events as a function of local time by using a local-time dependent value of D_c chosen so there are an equal number of events above D_c for each MLT. Fig. 3(b) shows that there is still a large relative difference between R_F in the noon and the midnight sectors and thus the difference in R_F between the two regions is not due to a difference in the number of events for a given threshold value.

III. FAILURE MODE ANALYSIS

A first step in determining the reason for model failure is to separate the time series by forecast outcome. One unexpected result of the analysis in Section II is that in the noon sector, $R_F(|dB_x/dt|)$ was much smaller than in midnight sector (Fig. 3). This is in contrast to the fact that these two sectors had approximately the same $PE(|dB_x/dt|)$ [Fig. 1(b)].

To determine possible reasons for this, we have taken the $|dB_x/dt|$ time series around the time of event forecasts and created averages for each forecast outcome. Fig. 4 shows these

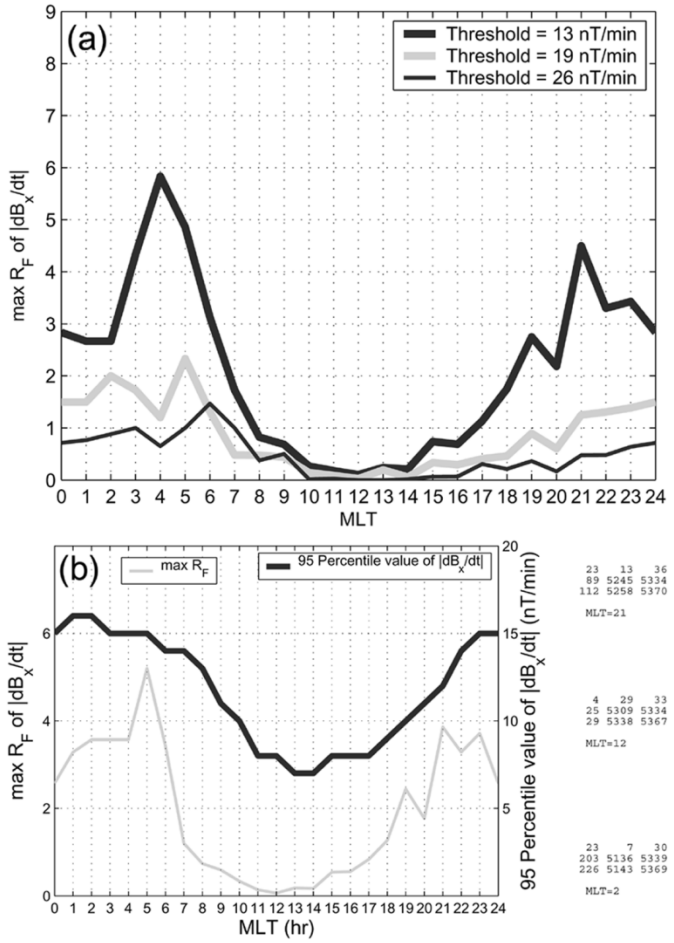

Fig. 3. (a) Ratio of the maximum (with respect to S_c) number of correct to false alarm forecasts of a threshold crossing in $|dB_x/dt|$ above threshold values corresponding to the 96, 98, and 99th percentiles of the distribution at that local time. (b) Ratio when the threshold value is chosen to be the 95th percentile value at a given MLT. To the right of the figures, contingency tables in the format of Table I are shown for selected local times.

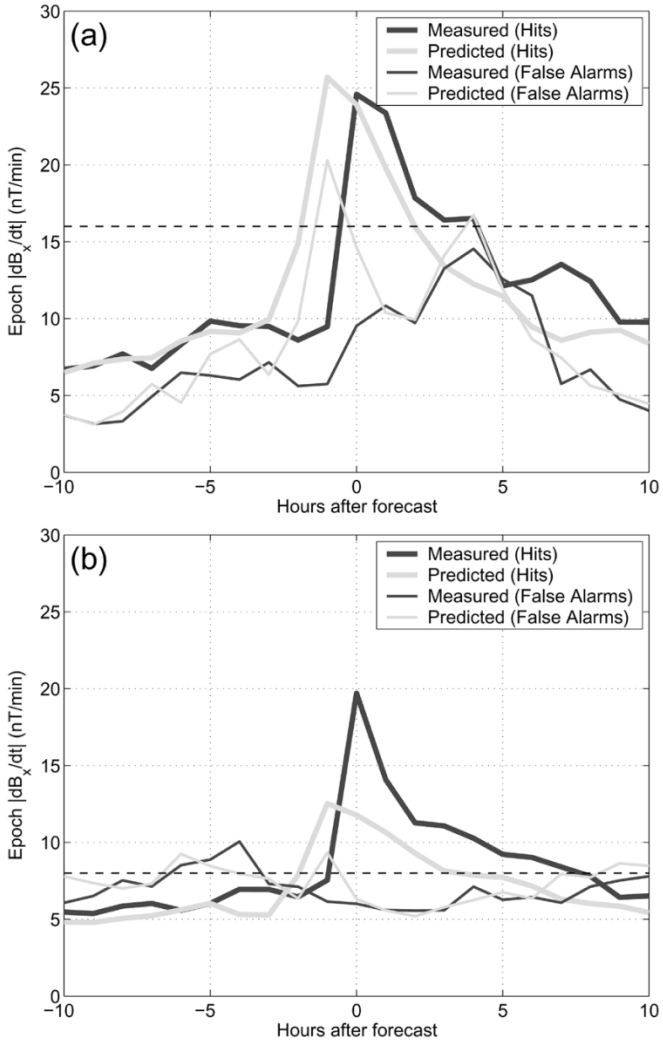

Fig. 4. Epoch averages of $|dB_x/dt|$ for forecast outcomes of hit and false alarm. (a) At $\mathrm{MLT} = 2$, the epoch averages are similar prior to the event, while at (b) $\mathrm{MLT} = 12$ events that were missed predictions of threshold crossings were pre-elevated prior to the actual crossing. Threshold values are shown as horizontal dashed lines.

epoch averages for D_c corresponding to the 95th percentile value at a given MLT. For the averages around the post-midnight sector $(\mathrm{MLT} = 2)$, the model was highly successful, with 23 hits and 7 false alarms, as was listed in the bottom contingency table of Fig. 3(b). Fig. 4(a) also shows that the epoch average of the prediction time series for both hits and false alarms is similar prior $t = 0$. Note that prior to the event, the predicted and hit time series are far below the threshold level. After $t = 0$, the false alarm prediction time series has a second peak and there is a corresponding rise in the measured time series; it appears that in these cases, the model is missing a factor that resulted in the initial suppression of a response.

Fig. 4(b) shows the epoch averages for D_c corresponding to the 95th percentile value of $|dB_x/dt|$ at $\mathrm{MLT} = 12$. Centered at this MLT, the model was comparatively unsuccessful, with 4 hits and 29 false alarms, as was listed in the bottom contingency table of Fig. 3(b). Fig. 4(b) also shows that the epoch average of the prediction time series for both hits and false alarms. For the false alarms, the prediction signal only jumped slightly above the threshold value for a single time interval. In contrast to Fig. 4(a), prior to $t = 0$, the predictions and the measurements hover near the threshold value. From this, we conclude that the high rate of false alarms is in part a result of the disturbance not being clearly separated from the threshold level prior to the event. This was verified by looking at the epoch average for all events (hits + misses). This curve (not shown) has a similarly flat profile prior to $t = 0$.

IV. Summary and Conclusions

A data-derived neural network model, originally developed to have a high prediction efficiency when predicting values of B_x and $|dB_x/dt|$, was analyzed for its ability to predict only large changes in these quantities. The performance metric used for event prediction was R_F, ratio of the number of correct forecasts to the number of false alarms. It was found that events in B_x were best predicted in the locations of the eastward- and westward-electrojets. The profile of $R_F(B_x)$ was similar to that of $PE(B_x)$, with the exception that the peak value of R_F was lower in the eastward than westward-electrojet regions.

In contrast, analysis of the $|dB_x/dt|$ model showed that its MLT profile of $R_F(|dB_x/dt|)$ differed significantly from that of $PE(|dB_x/dt|)$. The values of R_F were shown to be the highest in the midnight sector. The noon sector had comparatively low values of R_F. Using epoch analysis, a possible reason for the difference was found to be a result of $|dB_x/dt|$ in the noon sector being both elevated and near the threshold value prior to an event. In contrast, events levels in the midnight sector tended to be more separated from the threshold level prior to the event.

ACKNOWLEDGMENT

The authors thank the institutes that maintain the IMAGE magnetometer array, the PI institute for IMAGE: the Finnish Meteorological Institute (http://www.geo.fmi.fi/image), and the Space Radiation Laboratory at Caltech which provided the online ACE solar-wind data.

REFERENCES

[1] R. S. Weigel, A. Klimas, and D. Vassiliadis, "Solar-wind coupling to and predictability of ground magnetic fields and their time derivatives," *J. Geophys. Res.*, vol. 108, 2003.

[2] R. Matthews, "Decision-theoretic limits on earthquake prediction," *Geophys. J. Int.*, vol. 131, pp. 526–529, 1997.

[3] A. Thomson, "Evaluating space weather forecasts of geomagnetic activity from a user perspective," *Geophys. Res. Lett.*, vol. 27, pp. 4049–4052, 2000.

[4] D. Wilks, "A skill score based on economic value for probability forecasts," *Metrol. Applicat.*, vol. 8, pp. 209–219, 2001.

[5] C. Smith, J. L'Heureux, N. Ness, M. Acuna, L. Burlaga, and J. Scheifele, "The ACE magnetic fields experiment," *Space Sci. Rev.*, vol. 86, pp. 613–632, 1988.

[6] D. McComas, S. Bame, P. Barker, W. Feldman, J. Phillips, P. Riley, and J. Griffee, "Solar wind electron proton alpha monitor (SWEPAM) for the advanced composition explorer," *Space Sci. Rev.*, vol. 86, pp. 563–612, 1998.

[7] R. S. Weigel, D. Vassiliadis, and A. Klimas, "Coupling of the solar wind to temporal fluctuations in ground magnetic fields," *Geophys. Res. Lett.*, vol. 29, no. 19, 2002.

[8] R. S. Weigel, A. Klimas, and D. Vassiliadis, "Precursor analysis and prediction of large-amplitude relativistic electron fluxes," *Space Weather*, vol. 1, no. 3, pp. 1014–1020, 2003.

Robert S. Weigel received the Ph.D. degree in physics from the University of Texas, Austin, in 2000.

He is with the Laboratory for Atmospheric and Space Physics, the University of Colorado, Boulder. His interests include plasma physics, space weather prediction, and dynamical systems theory.

Dr. Weigel is a member of the American Geophysical Union.

Daniel N. Baker received the Ph.D. degree in physics from the University of Iowa, Iowa City, in 1974.

He is a Professor of physics in the Department of Astronomy and Planetary Sciences and serves as the Director of the Laboratory for Atmospheric and Space Physics Research at the University of Colorado, Boulder. His research interests include space instrument design and calibration, space physics data analysis, magnetospheric modeling, and public outreach to the space technology community and general public.

Dr. Baker is a member of the American Physical Society and the American Geophysical Union.

E. Joshua Rigler received the M.S. degree in aerospace engineering from the University of Colorado, Boulder, in 2000. He is currently working toward the Ph.D. degree in aerospace engineering with a focus on space weather forecasting using adaptive linear filters and state-space prediction models with extended sequential estimators (i.e., extended Kalman filters) at the University of Colorado, Boulder.

Mr. Rigler is a member of the American Geophysical Union.

Dimitris Vassiliadis received the Ph.D. degree in physics from the University of Maryland, College Park, in 1992.

He is with NASA's Goddard Space Flight Center with Universities Space Research Association. His interests include plasma physics, space science, and systems theory. He has recently developed a dynamic spatially extended model for the energetic electron flux in earth's radiation belts.

Dr. Vassiliadis is a member of the American Physical Society and the American Geophysical Union.