A Multi-Hour-Ahead global geospace model using Gated Recurrent Unit (GRU) networks and SuperMAG data

Andong Hu[1], Luiz F Guedes dos Santos[4], Jannis. Teunissen[2], and Enrico Camporeale[1,3]

1. CIRES, University of Colorado, Boulder, CO, USA
2. Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
3. NOAA Space Weather Prediction Center, Boulder, CO, USA
4. Shell USA Inc, San Benito, TX, USA

Correspondence: andong.hu@colorado.edu

1 Introduction

Geomagnetically induced currents (GICs) is one of the most severe risks posed by space weather events on ground and LEO infrastructures such as high-voltage power transmission systems. GICs are caused by sudden variations of the Earth's magnetic field that, through Faraday's law, induce an electric field. Hence, much attention has been dedicated to understanding and forecasting magnetic field perturbations ($\overrightarrow{B}$). In this study, we are aiming for modeling $\overrightarrow{B}$ using SuperMag data and machine learning (ML) approaches.

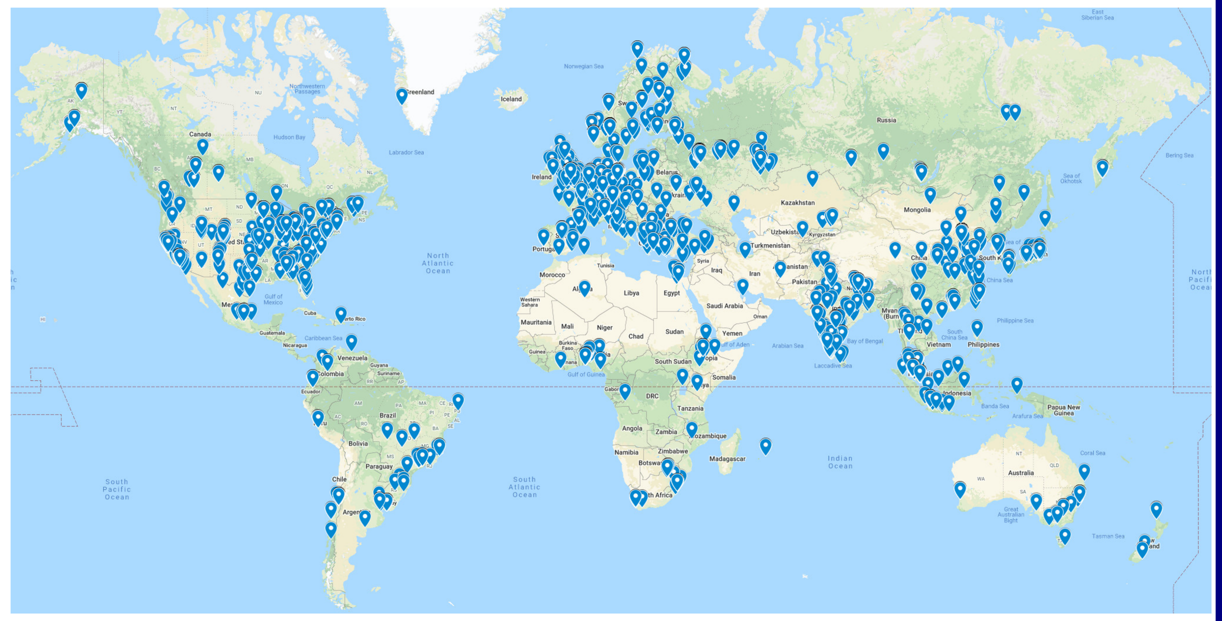


Figure 1: This figure shows location of the world's ground based magnetometers (blue dots). Notice the vast number of stations providing a powerful data set for global and continuous monitoring of the ground magnetic field.

2 Data and Methods

2.1 SuperMag

Overall, $\overrightarrow{B}$ measurements from 573 ground magnetometers during the 23rd solar cycle, i.e., 2000-2009 are downloaded and preprocessed via https://supermag.jhuapl.edu/mag/. This is because the solar activities during 23rd are more significant than those in 24th. The $\overrightarrow{B}$ measurements which is $GeoB$ after baseline removal, are used as the target of this study.

2.2 Omni Data

OmniWeb is one of the databases created by NASA SPDF which can provided solar wind data measured at L1 point (although some of the data have been propagated to the bow shock).
According Weimer (2013), several parameters are used for forming independent variables for this study:

- IMF B_T: the magnitude of the tangential IMF in the GSM Y-Z plane.
- θ_c: clock angle $\arctan(\frac{IMFB_y}{IMFB_z})$ (-90 to 90).
- V_{SW}: solar wind velocity.
- TA: dipole tilt angle expressed in radians, estimated as a function of DOY and UTC.
- $F_{10.7}$: The F10.7 index represents ionosphere conductivity variations due to solar ultraviolet radiation and is expressed as solar flux units (sfu).

Eventually we form the whole independent variable set (X) by: magnetic local time (MLT); declination (dec); $\overrightarrow{B^t_{n,e,z}}$; $F_{10.7}$; V_{SW}; TA; IMF B_T; $V_{SW}\times\sin\theta_c$; $t\times\sin\theta_c$; IMF $B_T\times\sin\theta_c$;
$\overrightarrow{B^{t+1}_{n,e,z}}$ is used as the target/response data (Y). Each $X(time-24:time)$ is corresponding to $Y(time+forecasthorizon)$ as one ML sample, where forecast horizon is how many hours ahead we would like to predict.

2.3 Storm Periods

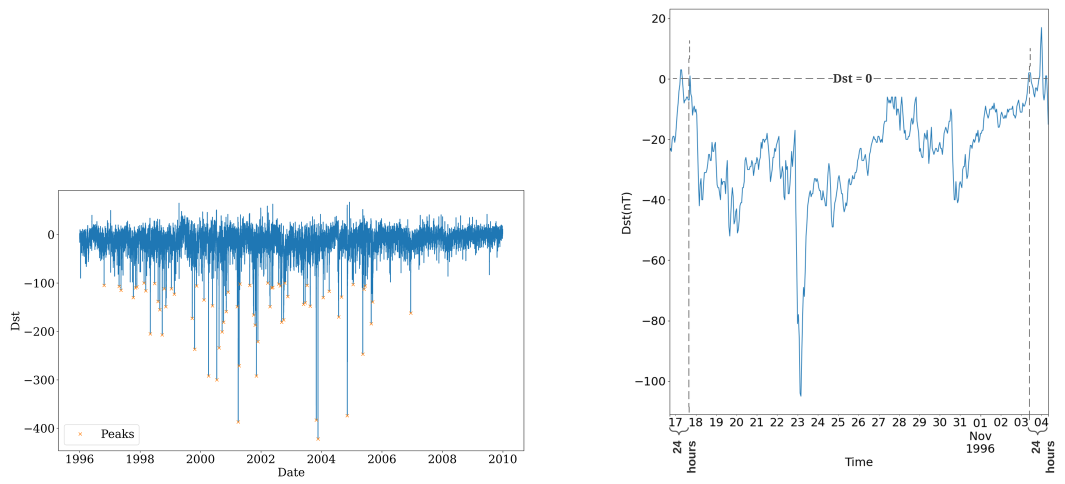


Figure 2: Left panel shows the time history of Dst during 1996 to 2010. X axis is date and Y axis is Dst value. The orange crosses denote peak values smaller than -100 nT, used for defining storm events considered in this study. Right panel is an example of the selection criterion to define the time range for one storm event. The Dst peak occurs on Oct. 23, 1996. The nearest positive Dst values before and after the peak occur on Oct. 18 and Nov. 03, respectively. The whole storm range is defined between Oct. 17, 1996 and Nov. 04, 1996 with a 24-hour buffer zone.

A storm period usually includes a pre-storm period, a main phase and a recovery phase. Define a storm period is always a difficult task. In this study, we look for the the nearest positive Dst values before and after each peak, and then extend the time window by a 24-hour buffer. An example is shown in the right panel of Fig. 2. With this procedure we make sure that the time intervals are selected in such a way that the negative Dst peaks do not always occur at the same time within the chosen storm-time window, hence the neural network does not memorize. All ML samples during one storm are considered as one event. Overall, the whole ML-ready data set includes 51 events for each given station as shown in the left panel of Fig. 2.

2.4 Method

Gated Recurrent Unit (GRU) Recurrent Neural Networks is used to give a preliminary prediction of from the ML-ready data set. The uncertainty of the model, so-called $\Delta\overrightarrow{B}$ model is then developed by the ACCURE method (Camporeale et al. (2021)). Then a linear estimator is then implemented for assimilating the $\Delta\overrightarrow{B}$ model into $\overrightarrow{B}$ model.

2.4.1 Gated Recurrent Unit (GRU) Recurrent Neural Networks

Gated Recurrent Unit (GRU) networks is one of the most widely used Recurrent Neural Networks (RNNs) which inherits the advantages of RNN. Similar to Long short-term memory (LSTM), GRU was created as the solution to short-term memory. In most scenarios, the performance of GRU is on par with LSTM, but computationally more efficient because of a less complex structure. The architecture of GRU is shown in Fig. 4. The X is a time series with a 6-hr span. Y is the corresponding Dst with a fixed time forecast horizon, i.e., 1-6 hours. The model trained for 1h ahead can be used for the 2h ahead. Fig. 3 exhibits the $\overrightarrow{B_e}$ RMSE over more than 150 stations. This implies that the developed model performs good in low an d mid latitudes, but not high latitudes.

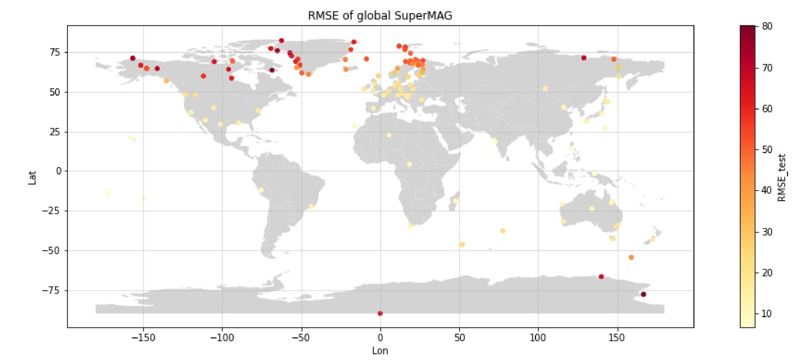
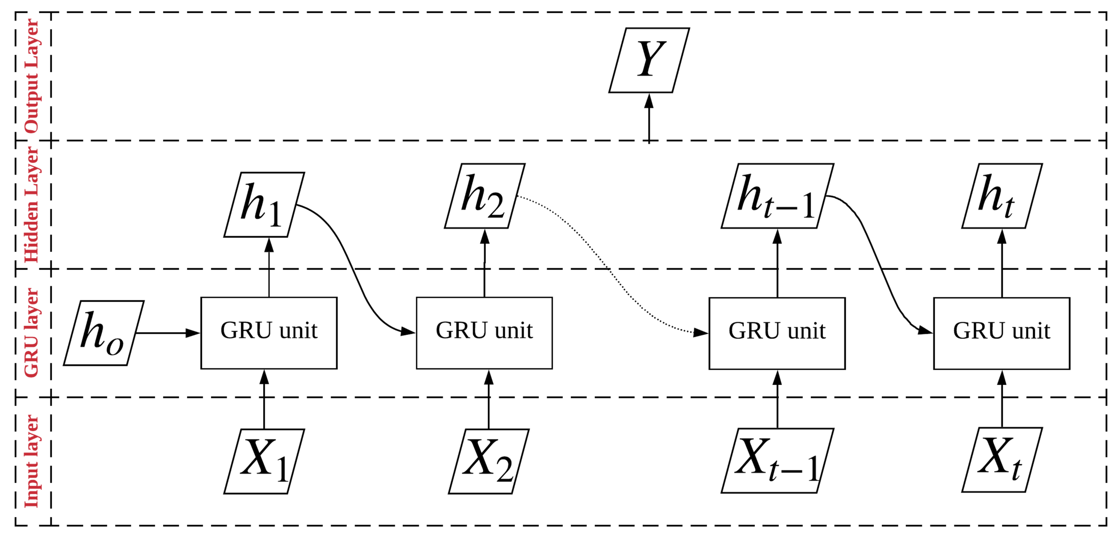


Figure 3: Global $\overrightarrow{B_e}$ RMSE of the developed GRU model from more than 150 SuperMAG stations.



$$z_t = \sigma_g(W_z x_t + U_z h_{t-1} + b_z)$$
$$r_t = \sigma_g(W_r x_t + U_r h_{t-1} + b_r)$$
$$\hat{h}_t = \Phi_h(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h)$$
$$h_t = (1-z_t)\odot h_{t-1} + z_t \odot \hat{h}_t$$

Figure 4: Structure of GRU. x_t is the independent variable set at the tth epoch, and Y is the target. h_t is the temporary results from the tth GRU unit, h_0 is manually initialized. The connection between h and Y are normal softmax/regression. Each GRU unit can be considered as a vanilla MLP model. z_t is update gate vector and r_t is reset gate vector. Ws, Us and bs are the coefficients that needed to be estimated during training. In addition, σg and Φh is sigmoid and tanh activation respectively.

2.4.2 Linear Estimator

Linear estimator is a simplified KF method. The equation is:

$$\hat{x}_i = \frac{\Delta Per^2}{\Delta\overrightarrow{B}^2_i + \Delta Per^2}\times x_i^{GRU} + \frac{\Delta Per^2}{\Delta\overrightarrow{B}^2_i + \Delta Per^2}\times x_i^{Per} \quad (1)$$

A general Kalman-filter is also used for comparison.
Generally, the whole procedures can be described into two steps:

- train $\overrightarrow{B}$ and $\Delta\overrightarrow{B}$ models for each forecast horizon (1-6 hrs ahead), as shown in the right panel of Fig. 4;
- assimilate $\overrightarrow{B}$ predictions from $\overrightarrow{B}$ model with $N-1$ hrs ahead into $\overrightarrow{B}$ predictions from $\overrightarrow{B}$ model with N hrs ahead to further improve the accuracy, as shown in Fig. 5.

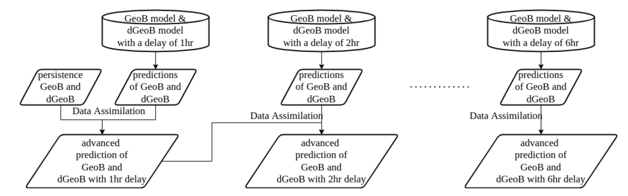


Figure 5: flowchart of data assimilation. E.g., for a given sample, we assimilate $\overrightarrow{B}$ model with $N-1$ hrs ahead into $\overrightarrow{B}$ predictions from $\overrightarrow{B}$ model with N hrs ahead to further improve the accuracy. When N=1, the $\overrightarrow{B}$ from the persistence model is assimilated for instead.

3 Results

An example of the final predictions of B_x for a given station in mid-latitudes during the 2003-Halloween storm are shown in Fig. 6. For each panel, yellow line is the GRU predictions; red line denotes real measurements; white line denotes the persistence model. It should be noted that the persistence model within forecast horizon:Nh denotes the model of forecast horizon:$N-1$h, except when $N=1$. Green line denotes the final results from DA. Blue and grey bars are the uncertainty (1 std) of GRU and persistence predictions respectively. It is clear that the DA results outperform GRU prediction significantly during the main phase of the storm.

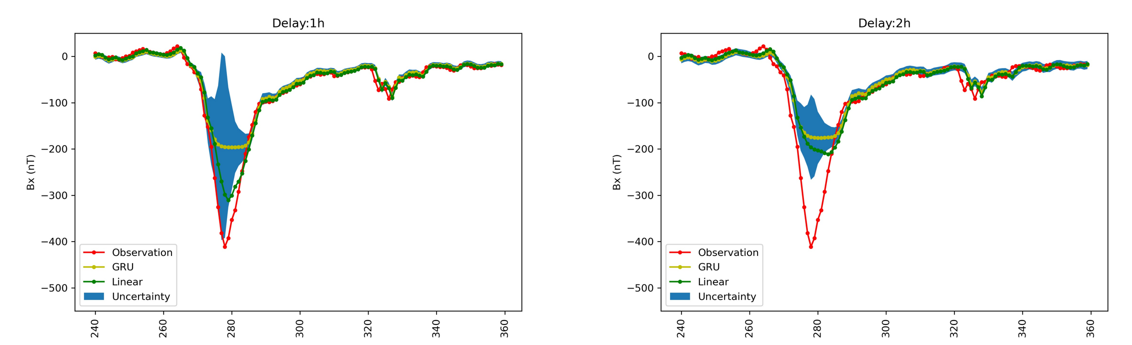


Figure 6: Final B_x predictions from a given station during the 2003 Halloween storm. It should be noted that the persistence model within forecast horizon:Nh denotes the model of forecast horizon:$N-1$h, except when $N=1$.

4 Summary and future

In summary, we wish to highlight the following:
○ The GRU and ACCRUE method can well predict $\overrightarrow{B}$ and the uncertainty of $\overrightarrow{B}$;
○ The $\overrightarrow{B}$ prediction can be significantly improved by assimilating the persistence model into the predictions.
○ A boost method will be implemented into this application in order to improve the performance of the model during main phase of strong storm periods

Acknowledgments

This research is funded by the National Aeronautics and Space Administration under grants 80NSSC20K1580, 80NSSC20K1275, 80NSSC21K1555. We thank OMNIWeb for providing the Dst data (https://omniweb.gsfc.nasa.gov/) and SuperMAG for providing magnetometer data (https://supermag.jhuapl.edu/).