Using a Convolutional Neural Network with Uncertainty to Forecast GIC Risk of Occurrence at Mid-Latitudes.

Mike Coughlan[1,2], Amy Keesee[1,2], Victor Pinto[1,2], Raman Mukundan[1,2], Hyunju Connor[3], Jeremiah Johnson[4]

[1]Department of Physics and Astronomy - University of New Hampshire
[2]Institute for the Study of Earth, Oceans, and Space - University of New Hampshire
[3]Department of Physics and Geophysical Institute - University of Alaska Fairbanks
[4]Department of Electrical and Computer Engineering - University of New Hampshire

MAGICIAN
Machine Learning Algorithms for Geomagnetically Induced Currents in Alaska and New Hampshire

Introduction

- The interaction between the solar wind and the Magnetosphere can produce Geomagnetically Induced Currents (GIC's) on the ground, which can cause power outages and damage to crucial infrastructure.
- The ability to predict when and where these events may occur could allow us to avoid the worst of this damage.
- The use of machine learning models can offer a computationally inexpensive method of predicting GIC events using horizontal dB/dt as a proxy, though most models thus far have fallen short of consistently accurate predictions. dB/dt was defined as:

$$\frac{dB}{dt} = \sqrt{\frac{dE^2}{dt} + \frac{dN^2}{dt}}$$

with N and E the North and East components of the magnetic field respectively.

- Here, the skill of a Convolutional Neural Network (CNN) was used to determine the risk of dB/dt going over a threshold of 7.15 nT/min for six ground magnetometer stations.
- The threshold was determined by taking the 99th percentile value of dB/dt for all magnetometer stations between 50 – 60 MLAT.
- Eight storms were chosen for testing and removed from the training set: March 30, 2001; August 31, 2001; May 15, 2005; August 31, 2005; December 16, 2006; April 5, 2010; August 5, 2011; March 17, 2015.
- The storms were based on the suite of GEM Challenge storms, minus the October 2003 storm due to the data gaps, with additional storms added from Welling, 2018.

Method/Data

- The model was trained exclusively on storm time data as defined by adding 12 hours of lead and 24 hours of recovery to SYM-H minimums of less than -50 nT .
- The storm data was extracted from a combined data frame of ACE and Supermag data.
- 6 Mid-latitude magnetometer stations were examined (MLAT averaged over period of 1995-2019): OTT (55.93º), NEW (54.92º), VIC (53.83º), WNG (50.29º), STJ (53.68º), and ESK (52.86º).
- The input features included solar wind speed (Vx, Vy, Vz), IMF_GSM (BT, By, Bz), proton density, dynamic pressure, SYM-H, AE INDEX, horizontal magnetic field (N,E), dB/dt, and ground magnetometer sin(MLT) and cos(MLT).
- The CNN model utilized 60 minutes of time history to determine if the dB/dt value would go above the 7.15 nT/min threshold, between 30 and 60 minutes into the future.
- The CNN models consists of 1 CNN layer, using 'RELU' activation. This was followed by a MaxPooling layer, a Flatten layer and two Dense layers with Dropout layers in between. The final layer is a Dense output layer using 'softmax' activation. The 'softmax' activation layer allows the model to interpret the inputs to the layer as discrete probability distributions, allowing us to interpret the outputs of a node as the probability of that node occurring. In this case the output of the node is the probability that dB/dt will cross the given threshold.
- The softmax activation can be described:

$$softmx(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}$$

- To determine the skill of the models a variety of metrics were calculated, including the Mean Bias and the Area Under the (Precision-Recall) Curve (AUC). The Root Mean Squared Error (RMSE) was also calculated, as was the standard deviation of the predicted values for comparison to the RMSE.
- The AUC metric utilizes a comparison of actual and predicted threshold crossings, where:
  - A is a True Positive, where both the actual and predicted cross the threshold
  - B is a False Positive, the actual does not cross but the predicted does
  - C is a False Negative, the actual crosses but the predicted does not
  - D is a True Negative, neither actual nor predicted cross the threshold
- The RMSE, as well as the Precision, and Recall metrics used for plotting the Precision-Recall curves, are defined as:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \qquad Precision = \frac{A}{A+B} \qquad Recall = \frac{A}{A+C} \qquad Bias = \frac{\sum_{i=1}^{n} \hat{y}_i}{n} - \frac{\sum_{i=1}^{n} y_i}{n}$$

- The Precision-Recall curve is calculated by changing the threshold above which the predicted data is considered positive, calculating the above Precision and Recall metrics, and then raising the threshold and repeating. The curve is then plotted with the resulting arrays.

Plots

Storm Plots

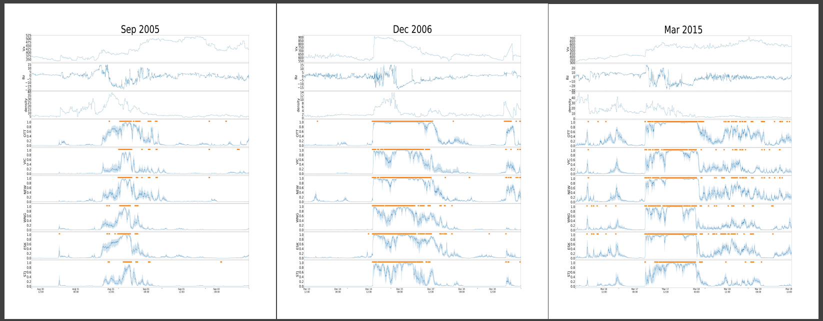

Figure 1: The real and predicted values for three selected storms. The orange bar at the top of the station panels indicates areas where the real data is positive, i.e. the dB/dt will be over 7.15 nT/min between 30 and 60 minutes in the future at the magnetometer station indicated on the left of the panel. The dark blue line indicates the mean probabilistic output of the positive node of the model. The light blue shaded area around the dark blue line indicates the 95% confidence interval calculated using 50 shuffled splits of the training data. The first three panels are solar wind Vx, Bz_GSM, and proton density at the L1 point as measured by ACE.

Precision-Recall Plots

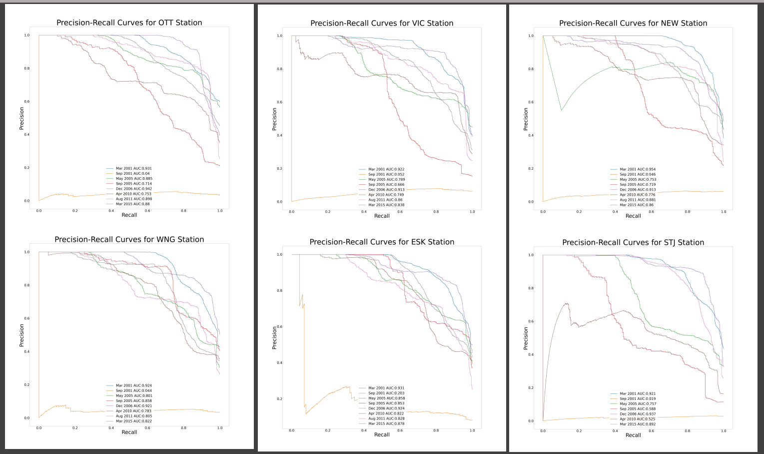

Figure 2: Precision-Recall curves for all stations and storms. Curves are drawn using the median result of the 50 shuffle split models using the calculated AUC as the indicator of the median. A perfect score for the AUC is a 1, and a score of 0 indicates a model that performs the same as a random chance model. The models perform well on all the storms with the exception of the September 2001 storm, the least intense storm in the testing suite, though the models still score better than random chance when testing on this storm for all stations.

Results

Metrics

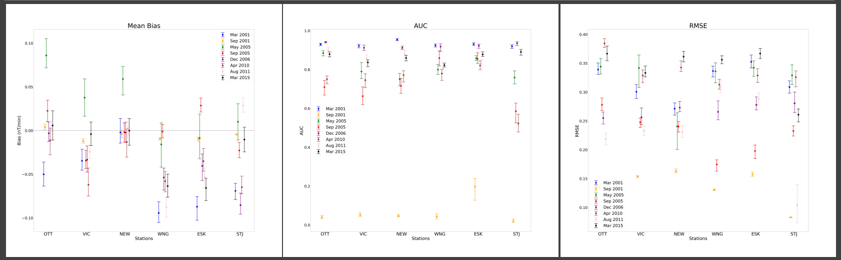

Figure 3: Mean Bias, AUC and RMSE results for all stations and storms. Mean values are plotted with the error bars indicating the 95% confidence interval. 1a) is the Mean Bias, where a positive score the mean results of the model are higher than the mean results of the real data indicating the model is over predicting, where a negative result indicates underpredicting. Our results show a trend of underprediction, though this metric is very broad. The AUC (3b) scores for most of the storms are above 0.6, indicating skillful models. The RMSE (3c) does not get above 0.4, though most of the results show room for improvement.

- The models appear to perform well, with relatively narrow confidence intervals, indicating a reliably designed model.
- Similar studies we have performed used data from the OMNI database, which is propagated to the bow shock. This study used ACE data from the L1 point. While a full study of the differences has not been performed, the AUC results in Figure 3b appear significantly higher than previous results. This is most likely due to the extra lead time provided by ACE being at the L1 point.
- The RMSE shows room for improvement given the nature of the 0,1 binary classification. The standard deviation of the predicted data was also calculated (not shown) to compare to the RMSE. The RMSE should be lower than the standard deviation, which was not universally the case in our results.
- The storm plots in Figure 1 show the models are capable of capturing the initial onsets of the spikes in dB/dt.
- Models show increases in probability when there are rapid changes in the solar wind conditions. For the Dec 2006 and Mar 2015 storms, there are clearly seen jumps in the solar wind velocity and proton density coinciding with the jump in the model probabilities. There appears to be a positive correlation between Bz variability and model probability output.
- Overall, while improvements can be made, the models show promise in their ability to forecast elevated levels of dB/dt with narrow confidence intervals, in a future time window using solar wind and ground magnetometer data.

Acknowledgements/References

This work was supported by NSF EPSCoR Award OIA-1920965

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.
- Camporeale, E., Cash, M. D., Singer, H. J., Balch, C. C., Huang, Z., & Toth, G. (2020). A gray-box model for a probabilistic estimate of regional ground magnetic perturbations: Enhancing the NOAA operational Geospace model with machine learning. *Journal of Geophysical Research: Space Physics*, *125*, e2019JA027684. https://doi.org/ 10.1029/2019JA027684
- Collado-Villaverde, A., Muñoz, P., & Cid, C. (2021). Deep Neural Networks with convolutional and LSTM layers for SYM-H and ASY-H forecasting. Space Weather, 19, e2021SW002748. doi: 10.1029/2021SW002748
- Gjerloev, J. W. (2012). The SuperMAG data processing technique. J. Geophys. Res. Space Phys. 117:A09213. doi: 10.1029/2012JA017683
- Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. Neural Comput. 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735
- Keesee AM, Pinto V, Coughlan M, Lennox C, Mahmud MS and Connor HK (2020) Comparison of Deep Learning Techniques to Model Connections Between Solar Wind and Ground Magnetic Perturbations. Front. Astron. Space Sci. 7:550874. doi: 10.3389/fspas.2020.550874
- Liemohn, M. W., Shane, A. D., Azari, A. R., Petersen, A. K., Swiger, B. M., & Mukhopadhyay, A. (2021). RMSE is not enough: Guidelines to robust data-model comparisons for magnetospheric physics. *Journal of Atmospheric and Solar-Terrestrial Physics*, *218*, 105624. doi:10.1016/j.jastp.2021.105624
- Lotz, S. I., and Cilliers, P. J. (2015). A solar wind-based model of geomagnetic field fluctuations at a mid-latitude station. Adv. Space Res. 55, 220–230. doi: 10.1016/j.asr.2014.09.014
- Pulkkinen, A., Rastätter, L., Kuznetsova, M., Singer, H., Balch, C., Weimer, D., 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
- Tóth, G., Meng, X., Gombosi, T. I., and Rastätter, L. (2014). Predicting the time derivative of local magnetic perturbations. J. Geophys. Res. Space Phys. 119, 310–321. doi: 10.1002/2013JA019456
- Welling, D. T., Ngwira, C. M., Opgenoorth, H., Haiducek, J. D., Savani, N. P., Morley, S. K., et al. (2018). Recommendations for next-generation ground magnetic perturbation validation. Space Weather 16, 1912–1920. doi: 10.1029/2018SW002064
- Wintoft, P., Wik, M., and Viljanen, A. (2015). Solar wind driven empirical forecast models of the time derivative of the ground magnetic field. J. Space Weather Space Clim. 5, A7–P9.