
Simultaneously forecasting global geomagnetic activity using recurrent networks

Charles Topliff

Georgia Institute of Technology
Atlanta, GA 30332
ctopliff0@gatech.edu

Morris Cohen

Georgia Institute of Technology
Atlanta, GA 30332
mcohen@gatech.edu

William Bristow

Pennsylvania State University
State College, PA 16801
wab5217@psu.edu

Abstract

Many systems used by society are extremely vulnerable to space weather events such as solar flares and geomagnetic storms which could potentially cause catastrophic damage. In recent years, many works have emerged to provide early warning to such systems by forecasting these events through some proxy, but these approaches have largely focused on a specific phenomenon. We present a sequence-to-sequence learning approach to the problem of forecasting global space weather conditions at an hourly resolution. This approach improves upon other work in this field by simultaneously forecasting several key proxies for geomagnetic activity up to 6 hours in advance. We demonstrate an improvement over the best currently known predictor of geomagnetic storms, and an improvement over a persistence baseline several hours in advance.

1 Introduction

Many ground systems are influenced or disrupted by extreme space weather, such as power infrastructure, satellite communication, television and radio links, as well as GPS navigation systems [Allen et al., 1989, Kan and Lee, 1979, Ding et al., 2007, Board et al., 2009]. Forecasts of space weather events would provide early warning to these systems, allowing for precautions to be taken to mitigate the effects of otherwise damaging activity. One area of particular interest is the modeling of variations in Earth’s magnetic field, or geomagnetic activity. There exist many different classes of geomagnetic activity, such as magnetic storms, magnetospheric substorms, quiet day variation, etc [Chapman and Bartels, 1962]. The physical processes behind many of these phenomena begin in the interior of the sun, where pressure builds to a release of a steady stream of radiation and energetic particles known as the solar wind. Geomagnetic storms at Earth are often the result of periods of drastically intensified solar wind known as coronal mass ejections. In recent years, the problem of forecasting such storms has gained much attention. Geomagnetic indices were developed to measure the general intensity of space weather from the ground Kivelson et al. [1995]. Recent works focus on modeling a particular class of geomagnetic activity by studying the relationship between solar wind measurements obtained from space weather datasets or derived from tertiary datasets, and a proxy for magnetic activity such as magnetic indices which summarize specific geomagnetic phenomena.

In this work, we seek to predict multiple magnetic indices such as the auroral electrojet indices (AE, AU, AL), the disturbance time index (Dst), the Solar Radio Flux at 10.7cm (F10.7) index, and the Planetary K (Kp) index several hours in advance. Where the majority of works have focused on

predicting a single magnetic index, we demonstrate an approach to simultaneously forecast several indices multiple hours in advance. In our work, we frame the problem as a multivariate sequence-to-sequence learning task [Sutskever et al., 2014] where the objective is to predict the four magnetic indices using solar wind measurements. We utilize long short-term memory networks which have been applied to myriad applications involving sequential data and are known for addressing vanishing gradient issues encountered with vanilla recurrent neural networks [Hochreiter and Schmidhuber, 1997].

2 Related Work

2.1 Magnetic Index Forecasting

Many works focus on modeling a particular class of geomagnetic activity by examining a single magnetic index. One example area of research is that of studying solar wind coupling where the objective is to compute some coupling function and solve a least-squares regression problem to predict a particular magnetic index using the time-lagged coupling function. McPherron et al. [2015] introduced a coupling function which partially linearizes the relationship between a set of commonly used solar wind variables and solved a least-squares regression problem to predict the AL index one hour in advance. They reported an average coefficient of determination (R^2) of .67 over two-month periods spanning 1963-2014 which improves upon other solar wind coupling functions [Newell et al., 2007, Rostoker et al., 1972, Clauer and Banks, 1986].

Other works which utilize neural networks have emerged in recent years. Revallo et al. [2014] investigated forecasting the Dst index one hour in advance using neural networks, and Boberg et al. [2000] sought to predict the Kp index in real time using solar wind measurements. In Maimaiti et al. [2019], the authors derived a substorm classification dataset by applying the substorm identification criteria [Newell and Gjerloev, 2011] to the SuperMAG SML index [Gjerloev, 2012]. The forecasting approach involves training variations of (previously) state-of-the-art convolutional networks, such as ResNet [He et al., 2016] to predict the onset of a magnetospheric substorm within the next hour.

2.2 Solar Wind Forecasting

Another area of research is the prediction of Solar Wind measurements at L1. NASA’s solar dynamics observatory curates a database of ultraviolet images of the solar corona [Galvez et al., 2019] which has been used by many works to forecast measurements at L1. Works typically model complex coupling processes which occur between the Sun and the Earth’s magnetic field in order to predict solar wind speed (V) or interplanetary magnetic field strength (B). Upendran et al. [2020] used a CNN-LSTM architecture leveraging a GoogLeNet [Szegedy et al., 2015] as a feature extractor to forecast daily averages of the solar wind multiple days in advance. Chandorkar et al. [2019] proposed a new regression framework to model the varying time lag implicit to the coupling processes between the solar magnetic field and solar wind propagation around the heliosphere.

3 Data and Preprocessing

3.1 Solar wind and IMF Measurements

We use measurements of the solar wind speed V , proton density n , and the components of the interplanetary magnetic field (IMF) B_x, B_y, B_z . These measurements are taken by the ACE and WIND satellites located at L1. We also utilize historical values of the magnetic indices AE, AL, AU, Dst, F10.7, and Kp. This data is curated by NASA’s Space Physics Data Facility and is available through the OMNIWeb database¹. As additional features, we include calendar year, day of year, and hour of day signals. These signals are converted to sinusoids with periods 11 years (roughly equivalent to the average length of a solar cycle), 365 days, and 24 hours in order to account for cyclical variations in the data. All data available on the OMNIWeb database has been preprocessed and resampled to several different time resolutions. For our experiments, we utilize hourly resolution data.

3.2 SuperDARN Measurements

We also utilize measurements derived from the SuperDARN network of radars [Greenwald et al., 1995]. Specifically, we used the spherical harmonic fit technique developed by Ruohoniemi et al.

¹OMNIWeb data available at <https://omniweb.gsfc.nasa.gov/>

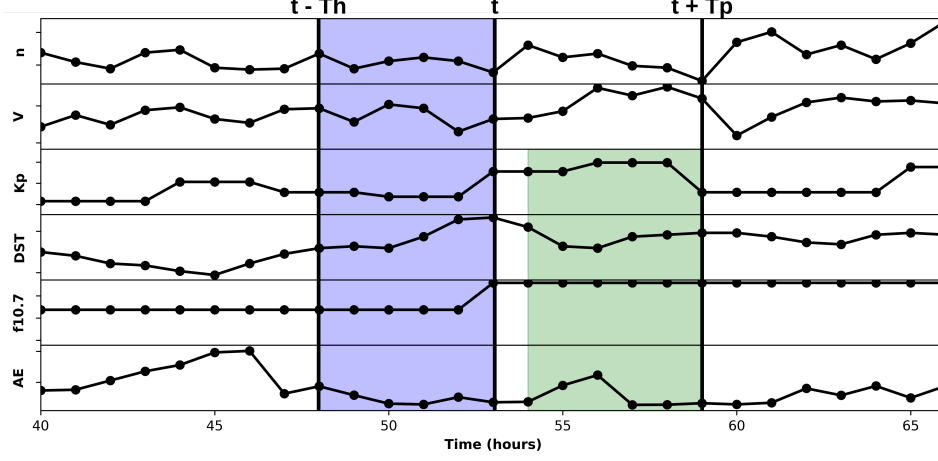


Figure 1: Structure of a single datum in our dataset. The blue shaded region between $t - T_h$ and t comprises all samples in the input time series and the green shaded region between t and $t + T_p$ comprises all samples in the target time series. In this example, $T_p = 6$, $T_h = 5$.

[1989] to produce a database of convection patterns covering the years from 2013 to 2017 (inclusive) on a five-minute cadence. These measurements were subsequently averaged and resampled to a 1 hour resolution. From these patterns we extracted two parameters, the Cross-Polar-Cap Potential (CCP), which parameterizes the degree to which solar wind energy is coupled into the Earth’s magnetosphere, (hence is one of the best parameters available for describing the level of global-scale magnetospheric activity), and the Polar Cap Radius (PCR), which is the average colatitude of the Convection Reversal Boundary. These measurements provide the best proxy for magnetic conditions at the poles aside from the auroral electrojet indices themselves. For a more detailed discussion, the reader is referred to [Bristow and Jensen, 2007].

4 Model Training

4.1 Dataset Assembly

We treat this forecasting problem as a multivariate sequence-to-sequence learning problem. At some time t , the inputs to our model comprise the samples of the time series of all input features in the interval $[t - T_h, t]$ where $T_h \geq 0$ is the number of samples in the past considered. The target is the samples of the time series of the four magnetic indices in the interval $(t, t + T_p]$ where $T_p \geq 1$ is the number of samples in the future predicted. Figure 1 illustrates the structure of input/output pairs on a subset of features. To assemble our datasets, we simply slide this structure through the entire time series one sample at a time. For all of our experiments, we discard all datapoints with missing values. To avoid violating causality, we split the data into training, validation, and testing sets sequentially with sizes being roughly 60, 10, and 30 percent, respectively. We center and scale the input features and training labels by the mean and standard deviation of each individual feature in the training set. We perform no additional preprocessing or filtering by the magnitude of the magnetic indices.

4.2 Model Training and Hyperparameter Tuning

In this work, we consider a simple neural network consisting of a LSTM which embeds the input time series into a low dimensional space followed by a linear layer which performs a linear transformation on the final output hidden state of the LSTM to the dimensionality of the output time series. We use Mean Squared Error as our loss function, and optimize the weights of the network using the Adam optimizer [Kingma and Ba, 2014] with weight decay. We conduct all experiments using the PyTorch deep learning framework [Paszke et al., 2019]. We treat the optimizer learning rate, weight decay coefficient, batch size, hidden dimension, and the number of layers of the LSTM as training hyperparameters. For each experiment, we conduct a random search [Bergstra and Bengio, 2012] over 100 total models and train the models for 1350 epochs using the median early stopping rule. We select the best model ranked by the performance on the validation set for all experiments.

5 Results

20 Year OMNIWeb Data In this experiment, we use 20 contiguous years of OMNIWeb Data (without including SuperDARN data) ranging from the years 2000-2019. We use a time history and a lead time of $T_h = T_p = 6$ hours. Our training, validation, and testing datasets correspond roughly to the years 2000-2011, 2012-2013, and 2014-2019, respectively. Our baseline for these experiments is a persistence forecast in which the prediction for the magnetic indices for all t is equivalent to the most recent available observation in the historical indices. In this experiment, we report performance in terms of pearson correlation coefficient ρ computed over the testing set. Table 1 illustrates the performance drop off over time for our network (mNet) and a persistence forecast (pers). The purpose of this experiment is to compare to a baseline persistent forecast, but it should be noted that we also compare to [McPherron et al., 2015] by computing the coefficient of determination (or explained variance) for a 1 hour AL forecast using predictions from the same network. We find that we achieve an R^2 of .744 over our test set ranging from 2014-2019. For this experiment, the best performing model uses a one layer LSTM, a batch size of 502, hidden dimension of 58, an initial learning rate of .00368, and a weight decay parameter of 9.7E-8.

Table 1: Pearson correlation coefficient for mNet and baseline (pers) predictions with test labels.

	AE		AU		AL		Dst		f10.7		Kp	
Hrs	mNet	pers	mNet	pers	mNet	pers	mNet	pers	mNet	pers	mNet	pers
1	.901	.478	.885	.491	.874	.422	.976	.790	.995	.993	.937	.658
2	.782	.452	.789	.462	.743	.397	.949	.759	.994	.991	.882	.632
3	.696	.429	.715	.437	.656	.375	.918	.730	.993	.990	.825	.605
4	.650	.406	.665	.414	.616	.352	.892	.703	.992	.989	.790	.578
5	.622	.382	.627	.394	.594	.328	.870	.678	.991	.988	.757	.557
6	.598	.364	.598	.379	.574	.309	.850	.655	.990	.987	.726	.535

SuperDARN measurements We compare the performance of our model on three different datasets to examine the added predictive capability of SuperDARN measurements. We start with the original OMNIWeb dataset as described in the previous experiment (base), create an additional dataset by replacing the historical values of the magnetic indices with the SuperDARN measurements (sdrn), and finally create a baseline dataset where we only use solar wind measurements (sw). The hyperparameters for the selected models are given in Table 2. For a fair comparison, we discard all datum corresponding to hours which are not represented in all three datasets. Similar to the first experiment, we use a time history and a lead time of $T_h = T_p = 6$ hours. Figure 2 illustrates the performance of the three models in terms of the pearson correlation coefficient of the predictions with test labels.

We find that the SuperDARN measurements alone do not provide a substantial improvement over the base solar wind model, and thus we cannot conclude on their efficacy as a sufficient proxy

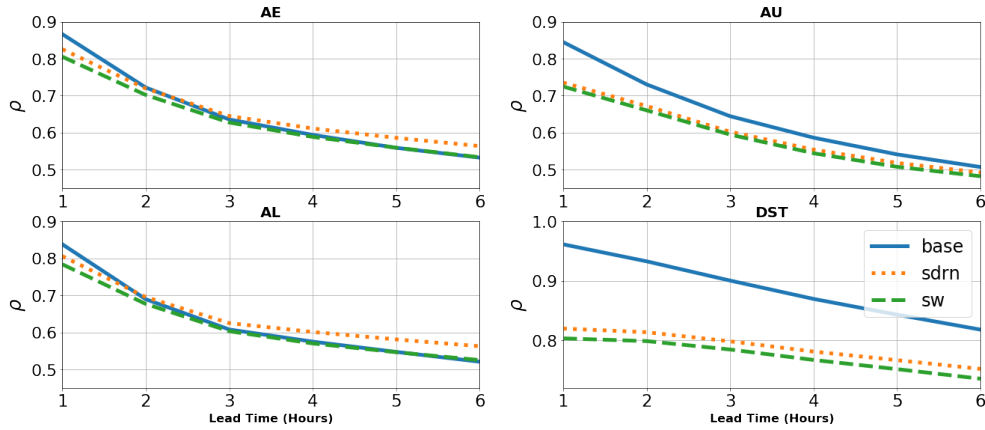


Figure 2: Pearson correlation for the predictions of the base model (blue/solid), sdrn model (orange/dotted), and sw model (red/dashed) with test labels.

Table 2: Hyperparameters for the models in the *SuperDARN measurements* experiment.

Model	LSTM Layers	Weight Decay	Hidden Dimension	Batch Size	Learning Rate
base	1	4.21E-8	59	1731	4.73E-2
sw	1	2.31E-7	44	1032	4.79E-2
sdrn	1	4.21E-2	48	1060	2.87E-6

for the magnetic indices. We also remark that the inconclusive results may be due to the limited data availability of SuperDARN measurements. To this end, we hope to increase the amount of SuperDARN measurements in future work.

6 Conclusion

We have presented a sequence-to-sequence learning approach to forecasting several geomagnetic indices simultaneously. These results demonstrate the efficacy of such an approach to forecasting these magnetic indices, an improvement over the current state-of-the-art for magnetospheric substorm prediction as well as an ability to outperform a persistence forecast over a longer time horizon. We find that the addition of the currently available SuperDARN data does not improve the predictive power of our models substantially. We seek to include additional results after more SuperDARN data has been processed in the future. Much additional future work is planned. For example, a larger model class may result in improved forecasts. Many recent works have emerged for similar time series forecasting problems including the use of attention, one-dimensional convolutional networks, and other recurrent neural networks. We seek to investigate these by performing a model architecture search [Elsken et al., 2018] or by using pretrained models similar to the approach used in [Maimaiti et al., 2019]. Additionally, prior works have used the ultraviolet coronal image dataset curated by NASA to forecast the solar wind and magnetic conditions at L1 orbit point at a coarse granularity. One potential future direction is to utilize a similar approach to forecast the magnetic indices directly and model the Earth-sun system in an end-to-end fashion. Such an approach may benefit from a pretrained convolutional network. Lastly, we seek to investigate the application of physics-informed neural networks to this problem by exploiting domain knowledge as a regularizer, such as the relationship between the AE, AL, and AU indices.

Broader Impact

Space weather is a broad term used to describe many potentially catastrophic phenomena for which we currently have limited forecasting capability. We know such events will happen, we know the consequences can be disastrous, but we do not have the sophistication to predict when or how a particular event will evolve. One example is a famous 1859 ‘Carrington Event’ that occurred during which northern lights were visible from Central America. Communication infrastructure was not nearly as advanced as it is today, but early telegraph stations caught on fire from induced electrical activity. The same thing could happen today to our electric power grid. The National Academies of Sciences estimates that an event of similar magnitude occurring today could cause several trillion dollars, knocking out power for many countries for many months. Currently, alerts to power utility companies about potential space weather disruptions are based exclusively on these geomagnetic indices, thus improving forecasts of these indices would immediately provide a tangible benefit to society. Space weather forecasting remains in its infancy, in part due to the cacophony of physical processes that cascade from the Sun to the Earth when a solar storm occurs. For decades, physicists have investigated this problem and have made substantial advances in modeling individual components of these complex systems, but have struggled to model the system with analytical techniques in an end-to-end fashion. As such, the broad impacts of effectively applying machine learning to the space weather forecasting problem could be huge, as ML techniques have a unique ability to describe systems that cannot be modeled analytically.

Acknowledgments and Disclosure of Funding

This work benefited directly from conversations with Jacob Bortnik and Ryan McGranaghan. We would like to thank them for their contributions.

This work was supported by the Defense Advanced Research Projects Agency (DARPA) under cooperative agreement D19AC00009 from the Department of the Interior to Georgia Tech. The first author is supported by the NDSEG fellowship.

References

- Joe Allen, Herb Sauer, Lou Frank, and Patricia Reiff. Effects of the march 1989 solar activity. *Eos, Transactions American Geophysical Union*, 70(46):1479–1488, 1989.
- JR Kan and LC Lee. Energy coupling function and solar wind-magnetosphere dynamo. *Geophysical Research Letters*, 6(7):577–580, 1979.
- F Ding, W Wan, B Ning, and M Wang. Large-scale traveling ionospheric disturbances observed by gps total electron content during the magnetic storm of 29–30 october 2003. *Journal of Geophysical Research: Space Physics*, 112(A6), 2007.
- Space Studies Board, National Research Council, et al. *Severe space weather events: Understanding societal and economic impacts: A workshop report*. National Academies Press, 2009.
- S Chapman and J Bartels. Geomagnetism, 2nd edn., vol. 1. *Clarendon, Oxford*, 1962.
- Margaret G Kivelson, Margaret Galland Kivelson, and Christopher T Russell. *Introduction to space physics*. Cambridge university press, 1995.
- Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.
- Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.
- Robert L McPherron, Tung-Shin Hsu, and Xiangning Chu. An optimum solar wind coupling function for the al index. *Journal of Geophysical Research: Space Physics*, 120(4):2494–2515, 2015.
- PT Newell, T Sotirelis, K Liou, C-I Meng, and FJ Rich. A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables. *Journal of Geophysical Research: Space Physics*, 112(A1), 2007.
- Gordon Rostoker, Hing-Lan Lam, and William D Hume. Response time of the magnetosphere to the interplanetary electric field. *Canadian Journal of Physics*, 50(6):544–547, 1972.
- C Robert Clauer and Peter M Banks. Relationship of the interplanetary electric field to the high-latitude ionospheric electric field and currents: Observations and model simulation. *Journal of Geophysical Research: Space Physics*, 91(A6):6959–6971, 1986.
- Miloš Revallo, Fridrich Valach, Pavel Hejda, and Josef Bochníček. A neural network dst index model driven by input time histories of the solar wind–magnetosphere interaction. *Journal of Atmospheric and Solar-Terrestrial Physics*, 110:9–14, 2014.
- Fredrik Boberg, Peter Wintoft, and Henrik Lundstedt. Real time kp predictions from solar wind data using neural networks. *Physics and Chemistry of the Earth, Part C: Solar, Terrestrial & Planetary Science*, 25(4):275–280, 2000.
- M Maimaiti, B Kunduri, JM Ruohoniemi, JBH Baker, and Leanna L House. A deep learning-based approach to forecast the onset of magnetic substorms. *Space Weather*, 17(11):1534–1552, 2019.
- PT Newell and JW Gjerloev. Substorm and magnetosphere characteristic scales inferred from the supermag auroral electrojet indices. *Journal of Geophysical Research: Space Physics*, 116(A12), 2011.
- JW Gjerloev. The supermag data processing technique. *Journal of Geophysical Research: Space Physics*, 117(A9), 2012.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016.
- Richard Galvez, David F Fouhey, Meng Jin, Alexandre Szenicer, Andrés Muñoz-Jaramillo, Mark CM Cheung, Paul J Wright, Monica G Bobra, Yang Liu, James Mason, et al. A machine-learning data set prepared from the nasa solar dynamics observatory mission. *The Astrophysical Journal Supplement Series*, 242(1):7, 2019.

- Vishal Upendran, Mark CM Cheung, Shravan Hanasoge, and Ganapathy Krishnamurthi. Solar wind prediction using deep learning. *Space Weather*, page e2020SW002478, 2020.
- Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- Mandar Chandorkar, Cyril Furtlehner, Bala Poduval, Enrico Camporeale, and Michele Sebag. Dynamic time lag regression: Predicting what & when. In *International Conference on Learning Representations*, 2019.
- RA Greenwald, KB Baker, JR Dudeney, M Pinnock, TB Jones, EC Thomas, J-P Villain, J-C Cerisier, C Senior, C Hanuise, et al. Darn/superdarn. *Space Science Reviews*, 71(1-4):761–796, 1995.
- JM Ruohoniemi, RA Greenwald, KB Baker, J-P Villain, C Hanuise, and J Kelly. Mapping high-latitude plasma convection with coherent hf radars. *Journal of Geophysical Research: Space Physics*, 94(A10):13463–13477, 1989.
- WA Bristow and Poul Jensen. A superposed epoch study of superdarn convection observations during substorms. *Journal of Geophysical Research: Space Physics*, 112(A6), 2007.
- Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. URL <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.
- James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, 13(1):281–305, 2012.
- Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. Neural architecture search: A survey. *arXiv preprint arXiv:1808.05377*, 2018.