Forecasting geomagnetic storms using long short-term memory neural networks

J. Peeperkorn, R. Dupuis, G. Lapenta

University of Leuven

Abstract

The goal of this research is to forecast the Disturbance Storm Time index (or Dst) 3-6 hours in advance. Different models have been used over the years to forecast the Dst and in recent years the use of machine learning, more specifically artificial neural networks, has made some promising results. The model used is a long short-term memory neural network, a variant of a recurrent artificial neural network designed to perform better on long term dependencies. The results of the model are evaluated using the root mean square error (RMSE) and Pearson's correlation coefficient (CC). A binary classification that divides the data into two categories (storm hour and no-storm hour) was also used. This was done with two different thresholds: one using the classification of a moderate storm and one using the classification of an intense storm. These classifications were then used in order to calculate a recall and precision rate of the storm hours forecasted.

Introduction

A **geomagnetic storm** is a disturbance in the Earth's geomagnetic field caused by an increase in the solar wind following for example a coronal mass ejection. The weakening of the Earth's magnetic field is measured by the **disturbance storm time** or Dst index. It is the average change in magnetic field based on the measurements of four low latitude stations. These disturbances can potentially cause big economical issues. For example in power grids it can lead to saturation of hard to replace transformers, resulting in possibly billions in economical damage. For mitigation purposes it can therefore be useful to forecast these geomagnetic disturbances multiple hours in advance.

Data

The input data of each forecast consists of a (multidimensional, multiple feature) **sequence of hourly data over multiple hours**. This sequence is then used to forecast the Dst index, 3-6 hours after the last data point in the sequence. The data used is provided by the OMNI 2 database of the National Space Science Data Center of NASA. The different features were selected using a correlation analysis, resulting in the **two different sets**:

1. Contains the Dst itself, the magnitude of the interplanetary magnetic field and its z-component and the solar wind parameters velocity, temperature and proton density. It uses data from 1998-2018.
2. Contains the Dst itself, the magnitude of the interplanetary magnetic field and its z-component, complemented with other geomagnetic indices. It uses data from 2004-2018.

Model

- An **artificial neural network**, build out of artificial neurons, is a supervised machine learning technique inspired by how the brain and it's biological neural network system solve problems.
- **Recurrent neural networks** (or RNN's) are designed to use information from different time steps taking into account the order of the sequence. Standard RNN's, that have a single activation layer in each time module, can not properly deal with long term dependencies.

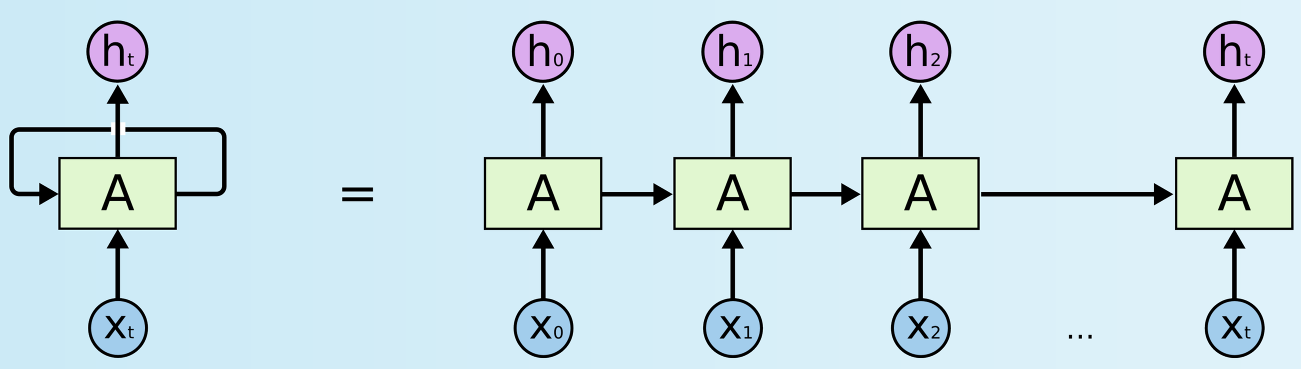


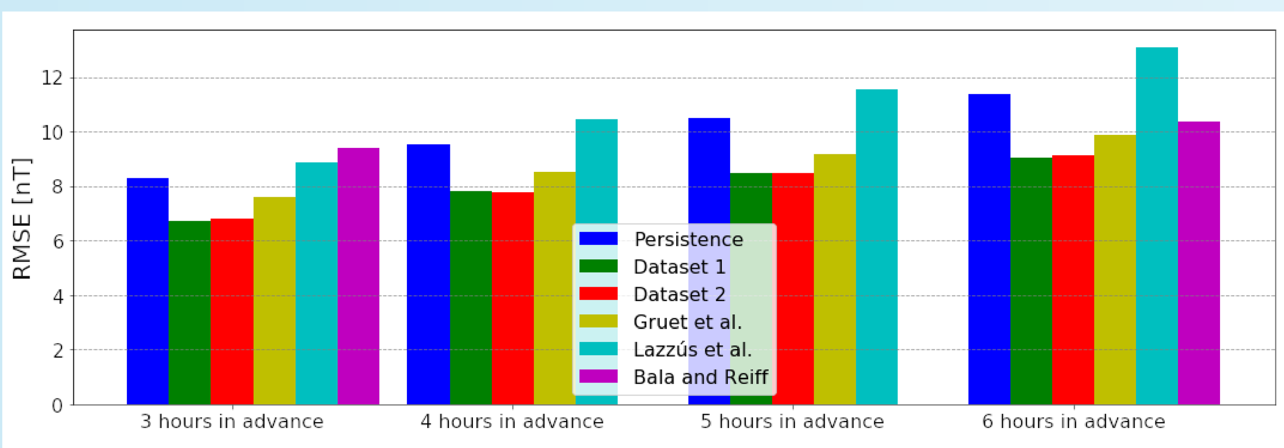
Figure 1: A depiction of an "unrolled" recurrent neural network. A RNN can been seen as a neural network where the hidden layer has a loop within it. Or when this loop is "rolled out", as multiple copies of a neural network passing through information. [6]

- **Long short-term memory** (or LSTM) neural networks are RNN's designed to deal with this issue. Each time module consists of four gates that can only change the message that is passed on with a linear operation, preserving the message passed on more properly.

The weights of the network were trained by using the Adam training algorithm, minimizing the mean squared error between the model's prediction and the real value of the Dst. [4] A hyperparameter grid search was done to fine tune the model, using HPC resources provided by VSC (Vlaams Supercomputer Centrum/Flemish Supercomputer Center).

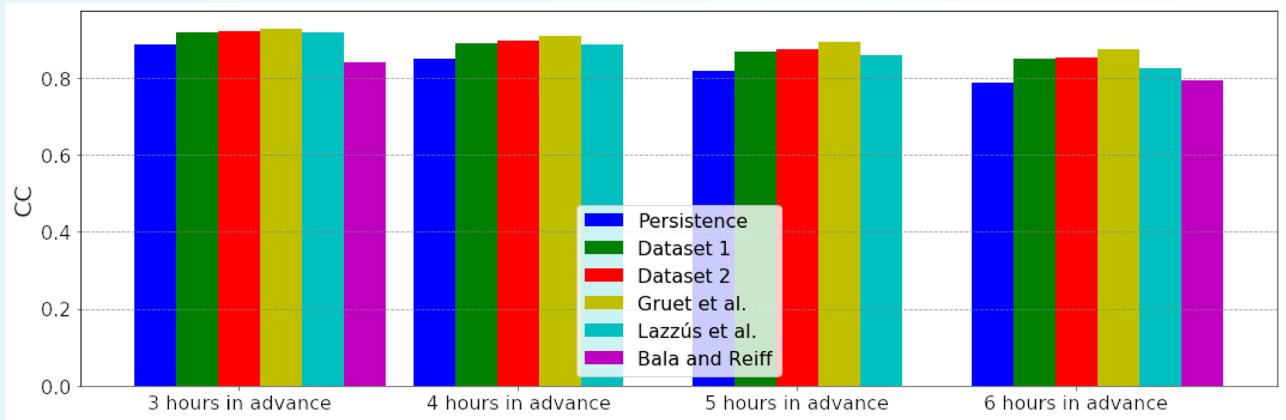
Results and evaluation

- The performance of the neural network is **compared** with other results from recent literature: Gruet et al. (2018) [2], Lazzús et al. (2017) [5] and Bala and Reiff (2012) [1].
- The model is also compared to a **persistence model** that just gives the last known value of the Dst as output.
- First two metrics used to evaluate results were **root mean squared** error (or RMSE) and **Pearson's correlation coefficient** (CC). Based on these two metrics the models using LSTM neural networks, ours an the one from Gruet et al. (2018), outperform the other models and **seem to perform well**.



	Persistence	Dataset 1	Dataset 2	Gruet et al.	Lazzús et al.	Bala & Reiff
t-3h	8.31 nT	**6.70 nT**	6.79 nT	7.59 nT	8.87 nT	9.40 nT
t-4h	9.53 nT	7.79 nT	**7.75 nT**	8.53 nT	10.44 nT	/
t-5h	10.49 nT	8.48 nT	**8.47 nT**	9.18 nT	11.56 nT	/
t-6h	11.35 nT	**9.04 nT**	9.11 nT	9.86 nT	13.09 nT	10.34 nT

Figure 2: The RMSE of the different models, forecasting the Dst different amount hours in the future.



	Persistence	Dataset 1	Dataset 2	Gruet et al.	Lazzús et al.	Bala & Reiff
t-3h	0.886	0.918	0.921	**0.928**	0.918	0.840
t-4h	0.851	0.890	0.896	**0.910**	0.887	/
t-5h	0.819	0.869	0.874	**0.892**	0.858	/
t-6h	0.788	0.848	0.854	**0.873**	0.826	0.795

Figure 3: The CC of the different models, forecasting the Dst different amount hours in the future.

- A **visual check** on some storm examples gives some more insight in how well an actual geomagnetic storms is forecasted by the network: it can be seen that the storms are actually often **forecasted too late**.

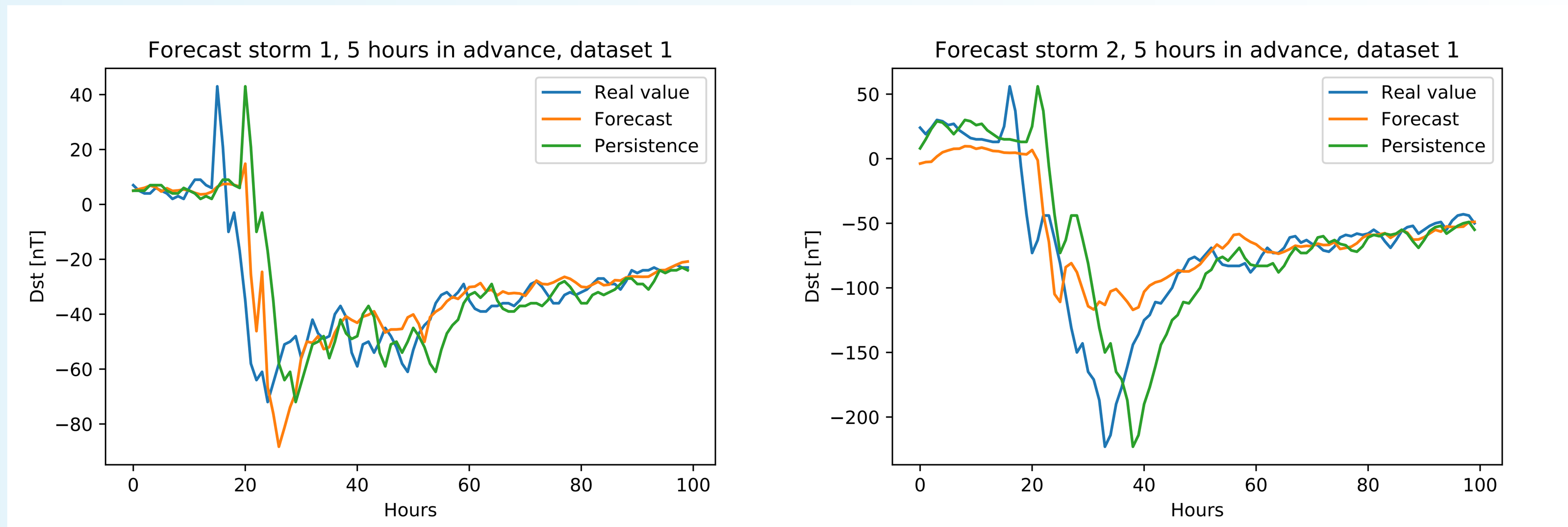


Figure 4: The forecast of the model and the persistence model on the moderate storm of 16/07/2017 and the intense storm of 17/03/2015.

- **Alternative metric** to measure the quality of the predictions takes two thresholds: -50 nT (moderate storms) and -100 nT (intense storm). The problem is converted into a binary classification problem on each storm hour individually and three rates are calculated:
  1. The **precision**: how many of the forecasted storm hours are actually storm hours)
  2. **recall** (how many of the storm hours are actually forecasted as storm hours)
  3. **f1** score: combining both
- The model has a relatively high precision rates, but **low recall rates**. Based on the recall rate, the model is even outperformed by the persistence model.
- The model is trained to be be act **reservedly** when forecasting storms.

	Model 1			**Model 2**			**Persistence**		
	Prec.	**Rec.**	**f1**	**Prec.**	**Rec.**	**f1**	**Prec.**	**Rec.**	**f1**
t - 3h	0.827	0.675	0.743	**0.837**	0.711	**0.769**	0.730	**0.724**	0.726
t - 4h	**0.806**	0.596	0.685	0.801	0.651	**0.718**	0.685	**0.680**	0.682
t - 5h	0.754	0.540	0.630	**0.773**	0.589	**0.669**	0.640	**0.635**	0.638
t - 6h	**0.737**	0.535	0.620	0.730	**0.592**	**0.654**	0.598	**0.592**	0.595

Table 1: The precision, recall and f1 scores of both models (two different datasets) and the persistence model based on the -50 nT threshold. The best values are indicated in bold.

	Model 1			**Model 2**			**Persistence**		
	Prec.	**Rec.**	**f1**	**Prec.**	**Rec.**	**f1**	**Prec.**	**Rec.**	**f1**
t - 3h	**0.840**	0.612	**0.708**	0.721	0.667	0.693	0.696	**0.688**	0.692
t - 4h	0.820	0.485	0.610	**0.862**	0.602	**0.709**	0.630	**0.624**	0.627
t - 5h	**0.844**	0.369	0.514	0.788	0.559	**0.654**	0.576	**0.570**	0.573
t - 6h	0.721	0.301	0.425	**0.917**	0.473	**0.624**	0.543	**0.538**	0.541

Table 2: The precision, recall and f1 scores of both models (two different datasets) and the persistence model based on the -100 nT threshold. The best values are indicated in bold.

Conclusions

- The LSTM model shows **potential** to be used in geomagnetic storms forecasting. The models obtained RMSE and CC comparable with the best result found in literature.
- A visual check revealed that the network was **not consistent in giving reliable forecasts**, not properly catching a lot of geomagnetic storm activity.
- Bad recall rates, obtained with the binary classification metric, indicate the bad handling of geomagnetic storms.
- Solely using RMSE and CC is **not sufficient** to confirm the quality of a forecast. This could mean that in other performed research the performance may also be overestimated.
- The high precision and low recall rates for the model also suggest that using a database with **more storm time** in relative to quiet time could improve the actual forecasting of geomagnetic storms, training the model to be less cautious.

References

[1] Ramkumar Bala and Patricia Reiff. Improvements in short-term forecasting of geomagnetic activity. *Space Weather*, 10(6), 2012.

[2] M. A. Gruet, M. Chandorkar, A. Sicard, and E. Camporeale. Multiple-hour-ahead forecast of the dst index using a combination of long short-term memory neural network and gaussian process. *Space Weather*, 16(11):1882–1896, 2018.

[3] J. H. King and N. E. Papitashvili. Solar wind spatial scales in and comparisons of hourly wind and ace plasma and magnetic field data. *Journal of Geophysical Research: Space Physics*, 110(A2), 2005.

[4] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv e-prints*, page arXiv:1412.6980, Dec 2014.

[5] JA Lazzús, CH López-Caraballo, P Rojas, I Salfate, M Rivera, and L Palma-Chilla. Forecasting the dst index using a swarm-optimized neural network. *Space Weather*, 15(8):1068–1089, 2017.

[6] Christopher Olah. Understanding lstm networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/, 2015. [Online; accessed 8-January-2019].