Solar Storm Prediction

Disha Sardana, Sangeetha Srinivasa

December 2016

Abstract

In order to accurately predict the occurrence of solar storms, we exploit the NARX neural networks to predict hourly Dst index in advance. Solar wind parameters such as interplanetary magnetic field (IMF), z component of magnetic field B_z, solar wind velocity V, solar wind plasma number density n are chosen as input parameters. The solar wind data and geomagnetic Dst index data are chosen over the period of 2000-2015, covering 91 storms and solar cycle 23 & 24. Dst is predicted over different periods : 1 year, 5 years, 10 years and 16 years, covering the onset, maxima and recovery phase of the solar storm cycle. For prediction, three different training algorithms are used: Levenberg-Marquardt, Bayesian Regularization and Scaled Conjugate Gradient. Data is randomly divided into 3 parts for each algorithm, 70% as training set, 15% as validation set and 15% as test set. It is found that storms of different strengths are correctly predicted and the correlation between the predicted Dst index and the observed Dst index (2000-2016) is found out to be 0.98. A low MSE in the range of [7,12] and a high Regression R-value of about 0.98 has been acheived.

1 Introduction and Motivation

When an intense and long lasting interplanetary convective electric field leads through substantial energization in the magnetosphere-ionosphere system to an intense ring current which is stronger than the threshold of the quantifying storm time Dst (disturbance storm time) index, the time interval is defined as a geomagnetic storm (GMS). [1]

GMSs are usually classified by the Dst indices as intense storms (peak Dst $\leq$ 100 nT), moderate storms (100 nT $<$ peak Dst $<$ 50 nT) and weak storms (peak Dst $>$ 50 nT). [2] In terms of time sequence, a GMS can be described in three phases: the initial, the main and the recovery phase. The initial phase may be gradual, or be represented by an abrupt change in the Dst, called a sudden commencement. The main phase of a storm is said to begin when the Dst assumes negative values and ends when it reaches its minimum decrease. The recovery phase, usually the longest one, is characterized by the returning of Dst to its pre-sudden commencement values. During a GMS, the Sun and the magnetosphere are connected giving rise to several changes both in interplanetary space and terrestrial environment. All the perturbations during GMSs involve energy transfer from solar wind to magnetosphere-ionosphere system and modify plasma and magnetic field there. Efficiency of a process seems to depend on the southern component of magnetic field and on the solar wind speed i.e., on the dawn-dusk component of solar wind electric

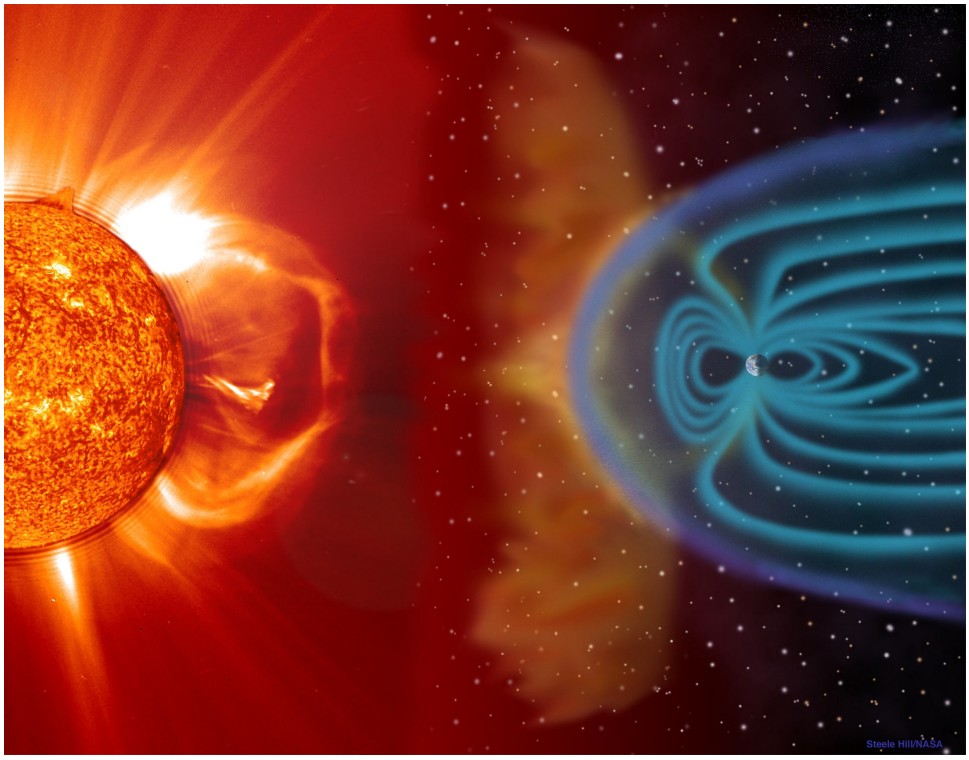

Figure 1: Solar storm hitting Earth

field. [2]

Dst, Kp, ap and AE indices are the four most commonly used geomagnetic indices (GI). Dst index is defined as the hourly average of the deviation of H (horizontal) component of magnetic field measured by several ground stations in mid to low latitudes and represents the degree of equatorial magnetic field deviation specifying the magnitude of GMSs. This is measured in the units of nano tesla (nT). Substorms, as one of the most important processes in the Earth's magnetosphere, receive wide attention in the space community because of their large scale influence on other magnetospheric and ionospheric dynamic processes in the global environment of the Earth. [3]

The study of magnetic storms is one of the main ingredients of space weather. These storms generate several changes both in interplanetary space and terrestrial environment and can damage the power supply, radio communications and spacecrafts. In order to understand the response of the magnetosphere to interplanetary conditions during GMSs, several studies have derived relations between interplanetary values of various plasma and field parameters and various combinations of these parameters. A unique relationship which may ultimately lead to an unambiguous understanding of the phenomenon and predict the occurrence of GMSs is yet to emerge.

In our project, results are presented for the relationship between Dst and several plasma and field parameters for a 91 GMSs selection of solar cycle 23 and 24. In Section 2 & 3 data sources, selection criteria of data set and visualization of data have been

presented. In the following sections, model building algorithms, error plots, results with correlation between GI and plasma and field parameters, analyses and study of four different GMSs periods: 1 year(2011), 5 years(2011-2015) and 10 years(2006-2015), 16 years(2000-2015) covering the rise, maximum, and decay phases, have been investigated. Discussion of results and final conclusions are presented in the later sections.

2 Data Description

Conditions in the solar wind resulting in magnetic storms on the Earth have been a subject of long and intensive investigation. The solar wind plasma and field measurements with 1 hour time resolution were obtained from the OMNI website: **http://omniweb.gsfc.nasa.gov**. Hourly Dst indices were obtained from the world data center at the University of Kyoto database: **http://wdc.kugi.kyoto-u.ac.jp/dstdir/index.html**.

OMNI data center provide average magnitude of interplanetary magnetic field IMF(nT), negative z-component of IMF B_z (nT), solar wind plasma number density n (cm^{-3}) and the solar wind velocity V (km/s). All the field parameters are in Geocentric Solar Magnetospheric (GSM) coordinate system. They also provide proton temperature T (K), proton density D (cm^{-3}), flow pressure P (nPa), electric field E (mV/m), plasma beta (β), field magnitude average FMA (nT), variance of total IMF σB (nT) etc.

The 4 features that were selected as input features were based on their direct correlation with Dst index and how much it gets affected as these parameters change. [4]

3 Data Pre-processing

Input Values: The data from the OMNI data center needed no data pre-processing. There were no missing values or (####/ 99999) values. Normalization wasn't needed since the input features' values were used as such to predict the Dst indices. The data was in .lst format, so no format conversions were needed as such.

Target Values: However, the data from Kyoto database was given in .dat format. It had to be converted to .csv format to use it in MATLAB code. Also, it had about 8-10 lines of text data after every month of data which needed to be cleaned. Apart from the data cleaning, there weren't any missing values or (#### / 9999) values. Normalization wasn't needed here too.

4 Data Exploration and Statistics

Using the Kyoto database, storms were identified on the basis of Dst index values. Those having Dst indices less than -100 nT were identified as storms. Then these storms were classified on the basis of the year, month and their onset times using MATLAB. In total there were 91 storms and the whole solar cycle was covered. Figure 2 below represents the hourly Dst index over the period 2000-2015 and was plotted using Weka. Figure 3,4,5 are the bar graphs showing the storm counts.

Sr. No.	Statistic	Value
1.	Minimum	-422
2.	Maximum	95
3.	Mean	-12.285
4.	Standard Deviation	21

Table 1: Statistics 2000-2015

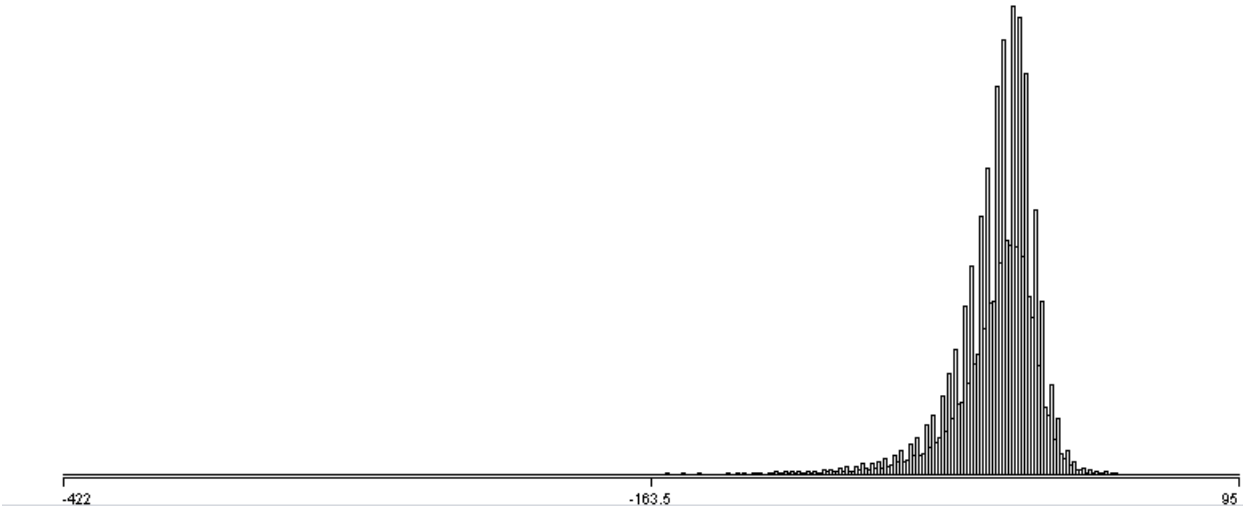

Figure 2: Hourly Dst index 2000-2015

Dst index < -100 nT

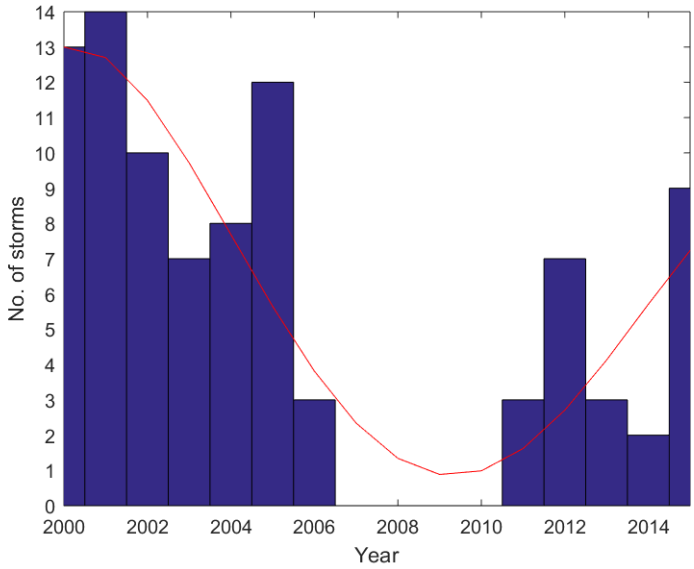

Figure 3: Classification of storms on the basis of year

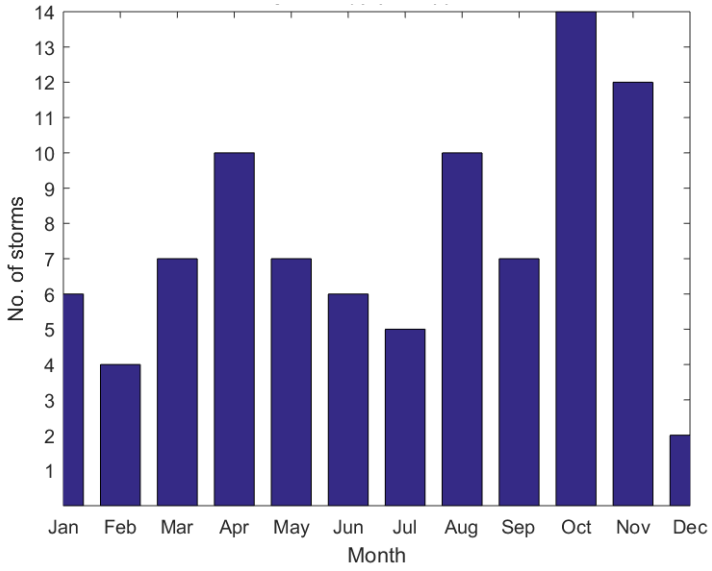

Figure 4: Classification of storms on the basis of month

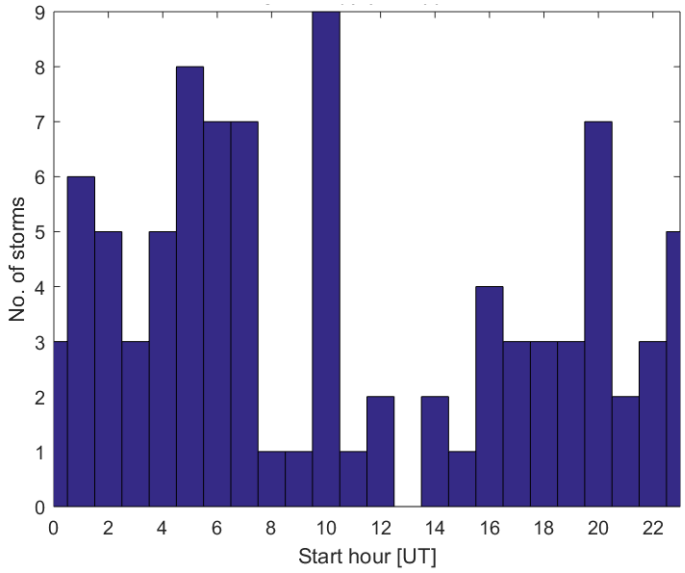

Figure 5: Classification of storms on the basis of onset time

The main idea was to identify a pattern on the month, year or onset times basis. It is observed that there are no storms from year 2007-2010, which ought to happen since that is a solar minimum phase.

Also, it is observed that maximum number of storms happened in the month of October.

5 Model Building

We describe below the NARX model and the three training algorithms that were used in our implementation.

5.1 NARX - Nonlinear autoregressive network with exogenous inputs

The nonlinear autoregressive network with exogenous inputs (NARX) is a recurrent dynamic network, with feedback connections enclosing several layers of the network. The NARX model is based on the linear ARX model, which is commonly used in time-series modeling.

The defining equation for the NARX model is

$$y(t) = f(y(t-1), y(t-2), \ldots, y(t-n_y), u(t-1), u(t-2), \ldots, u(t-n_u))$$

where the next value of the dependent output signal y(t) is regressed on previous values of the output signal and previous values of an independent (exogenous) input signal.

The standard NARX network is a two-layer feedforward network, with a sigmoid transfer function in the hidden layer and a linear transfer function in the output layer. This network also uses tapped delay lines to store previous values of the x(t) and y(t) sequences. Note that the output of the NARX network, y(t), is fed back to the input of the network (through delays), since y(t) is a function of y(t – 1), y(t – 2), ..., y(t – d).

This model could be used to predict future values of a stock or bond, based on such economic variables as unemployment rates, GDP, etc. It could also be used for system identification, in which models are developed to represent dynamic systems, such as chemical processes, manufacturing systems, robotics, aerospace vehicles, etc. [5]

5.2 Training Algorithms

5.2.1 Levenberg-Marquardt

Levenberg-Marquardt is a popular alternative to the Gauss-Newton method of finding the minimum of a function F(x) that is a sum of squares of nonlinear functions,

$$F(x) = \frac{1}{2}\sum_{i=1}^{m}[f_i(x)]^2.$$

Let the Jacobian of $f_i(x)$ be denoted $J_i(x)$, then the Levenberg-Marquardt method searches in the direction given by the solution p to the equations

$$(J_k^T J_k + \lambda_k I)p_k = -J_k^T f_k,$$

where λ_k are nonnegative scalars and I is the identity matrix. The method has the nice property that, for some scalar Δ related to λ_k, the vector p_k is the solution of the constrained subproblem of minimizing $||J_k p + f_k||_2^2/2$ subject to $||p||_2 <= \Delta$ [6].

5.2.2 Bayesian Regularization

Bayesian Regularization is a network training function that updates the weight and bias values according to Levenberg-Marquardt optimization. It minimizes a combination of squared errors and weights, and then determines the correct combination so as to produce a network that generalizes well.

Bayesian regularized artificial neural networks (BRANNs) are more robust than standard back-propagation nets and can reduce or eliminate the need for lengthy cross-validation. Bayesian regularization is a mathematical process that converts a nonlinear regression into a "well-posed" statistical problem in the manner of a ridge regression. The advantage of BRANNs is that the models are robust and the validation process, which scales as O(N2) in normal regression methods, such as back propagation, is unnecessary. [7]

They are difficult to overtrain, since evidence procedures provide an objective Bayesian criterion for stopping training. They are also difficult to overfit, because the BRANN calculates and trains on a number of effective network parameters or weights, effectively turning off those that are not relevant. This effective number is usually considerably smaller than the number of weights in a standard fully connected back-propagation neural net. [7]

5.2.3 Scaled Conjugate Gradient

SCG is a supervised learning algorithm for feedforward neural networks, and is a member of the class of conjugate gradient methods. [8] From an optimization point of view learning in a neural network is equivalent to minimizing a global error function, which is a multivariate function that depends on the weights in the network. Many of the training algorithms are based on the gradient descent algorithm.

Minimization is a local iterative process in which an approximation to the function in a neighborhood of the current point in the weight space is minimized. Most of the optimization functions used to minimize functions are based on the same strategy. The Scaled Conjugate Gradient algorithm denotes the quadratic approximation of the error E in the neighborhood of a point w by:

$$E_q w(y) = E(w) + E^{'}(w)^T y + \frac{1}{2} y^T E^{''}(w) y$$

In order to determine the minimum to $E_q w(y)$ the critical points for $E_q w(y)$ must be found. The critical points are the solution to the linear system defined by Moller:

$$E^{'}_{qw}(y) = E^{''}(w)y + E^{'}(w) = 0$$

SCG belongs to the class of Conjugate Gradient Methods, which show superlinear convergence on most problems. By using a step size scaling mechanism SCG avoids a time consuming linear search per learning iteration, which makes the algorithm faster than other second order algorithms. [9]

6 Model Evaluation

A comparative study against the three training algorithms explained in the Model Building Section of NARX model was done by varying:

(i) the number of neurons in the hidden layer
(ii) the data input to the model; our model was trained with four different datasets: 2000-2015, 2006-2015, 2011-2015 and 2011.
(iii) An exploratory study of the impact of having all features vs a subset of features on our classification model was carried out.

6.1 Subset feature selection

A comparison was done between Dst index values predicted from 10 input features versus 4 input features for the year 2011.

The 10 input features that were selected are: Interplanetary Magnetic Field IMF Avg Mag (nT), z component of IMF i.e. B_z (nT), Variance σ in IMF Avg. Mag. (nT), Longitude of Avg. IMF (deg), Variance σ in B_z (nT), Proton Temperature (K), Proton Density (n/cc), Flow Speed (km/s), Alpha/Proton Density Ratio, Flow Pressure (nPa)

Below is the regression plot between the predicted output and target values (Dst index) when 10 input features were selected. The closer the value of R to 1, the better is the prediction.

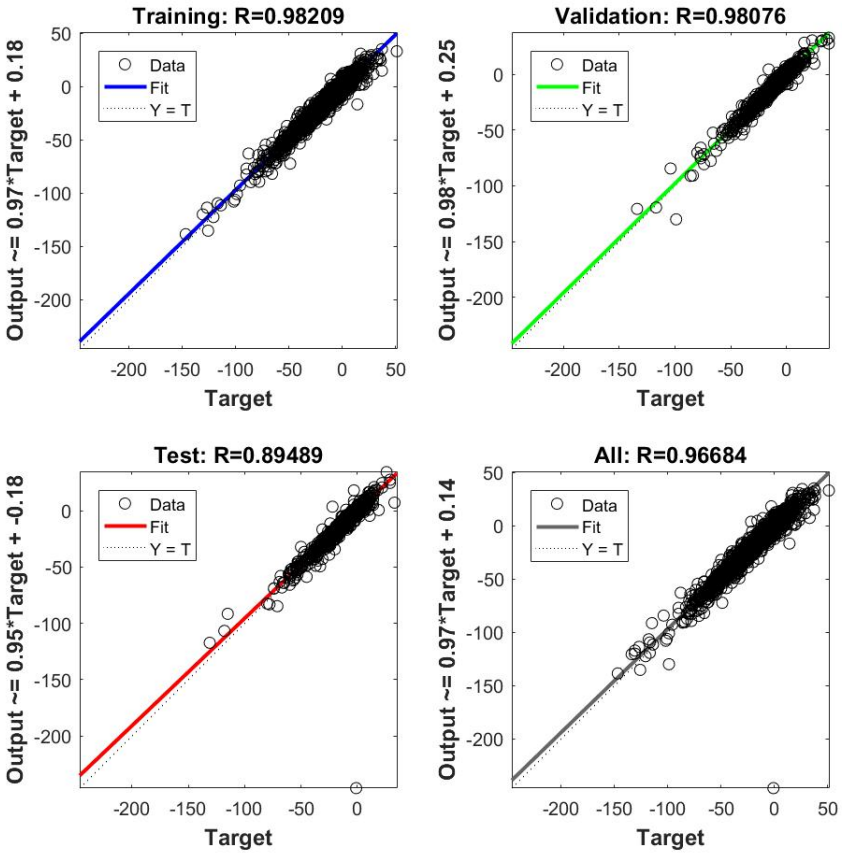

Figure 6: Regression plot when 10 features are given as input

The 4 input features that were selected are: Interplanetary Magnetic Field IMF Avg Mag (nT), z component of IMF i.e. B_z (nT), Proton Density (n/cc), Flow Speed (km/s)

These 4 input features have the strongest relation with the Dst index [4], hence we used these.

Below is the regression plot between the predicted output and target values (Dst index) when 4 input features were selected.

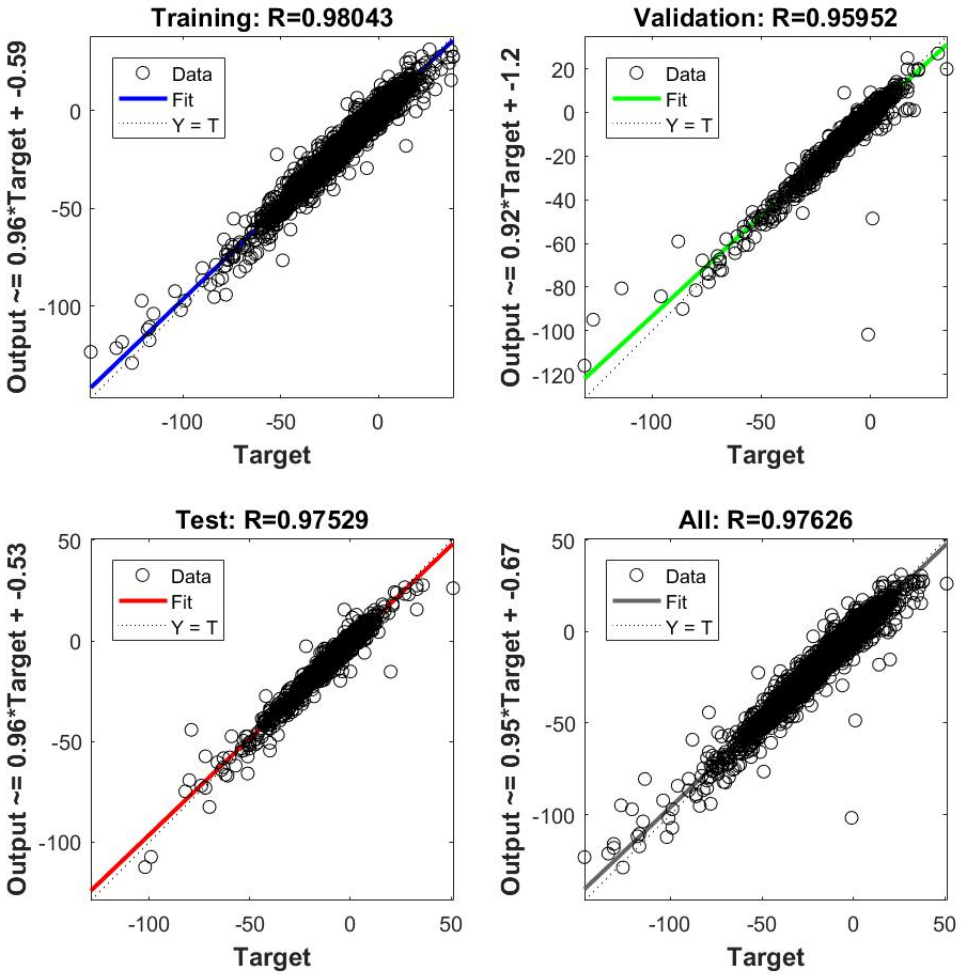

Figure 7: Regression plot when 4 features are given as input

It is observed that the R value of test has increased significantly when we take 4 input features as compared to when we take 10 input features from R=0.89489 to R=0.97529.

The overall R value also slightly improves as we choose 4 features instead of 10.

Also, the computation time and model complexity decreases with less number of input features. Hence, we used these 4 features.

6.2 MSE

Below are the results obtained on calculating MSE between the target data and the output test data.

Algorithm		MSE		
		2 neurons	4 neurons	8 neurons
Levenberg	2000 - 2015	12.82831	13.47398	12.47393
	2006 - 2015	8.06696	7.95993	8.66653
	2011 - 2015	11.18627	10.44852	9.98104
	2011	11.58896	14.68709	10.22823
Bayesian	2000 - 2015	12.45313	12.41938	11.44120
	2006 - 2015	8.33769	8.21845	8.09592
	2011 - 2015	10.55661	10.27141	9.99965
	2011	9.03188	10.30126	126.82591
SCG	2000 - 2015	15.10143	25.3449	31.30577
	2006 - 2015	9.12728	10.13344	13.17883
	2011 - 2015	41.54599	12.61923	15.08707
	2011	32.86807	11.16878	40.02773

Table 2: MSE of testing data

Mean Squared Error is the average squared difference between outputs and targets. Lower values are better. Zero means no error.

- As can be seen from the table, for Levenberg and Bayesian MSE decreases as the number of neurons in the hidden layer increases.
- On the contrary, SCG exhibits better performance with 2 neurons in hidden layer for 16 years and 10 years dataset whereas for 5 years and 1 year dataset 4 neurons in hidden layer gives better performance.

We present in Table 3 our findings of the best validation performance of the three algorithms on the different datasets.

We have divided our data such that 70% is training, 15% is validation and the remaining 15% is test data. This division is done randomly. Our training model continues to build the model till all the validation tests pass successfully. Our model is run for a maximum of 1000 iterations in order to achieve this. Table 3 presents our findings

Algorithm		Mean Square Error (MSE)			Epoch		
		2 neurons	4 neurons	8 neurons	2 neurons	4 neurons	8 neurons
Levenberg	2000 - 2015	12.8086	12.7301	11.652	105	89	119
	2006 - 2015	8.2044	8.8119	7.8432	66	102	78
	2011 - 2015	10.8794	10.294	11.0657	10	51	14
	2011	9.9019	11.3606	7.3647	35	12	17
Bayesian	2000 - 2015	12.2822	11.5602	11.0836	160	1000	479
	2006 - 2015	8.6496	7.956	7.8565	135	392	591
	2011 - 2015	10.7697	9.7836	9.6028	430	231	655
	2011	8.8668	8.4336	7.7692	78	359	390
SCG	2000 - 2015	15.2609	28.2192	35.1186	250	178	71
	2006 - 2015	18.7943	8.1442	8.3684	109	51	607
	2011 - 2015	24.6017	27.1234	13.4905	67	62	135
	2011	214.448	21.9646	65.5691	5	55	7

Table 3: Best Validation Performance - MSE

of MSE of the validation data. Epoch represents the number of iterations at which the validation tests passed and gave the best performance.
A pattern similar to MSE of test data is observed even here.

- For Levenberg and Bayesian, MSE decreases as the number of neurons in the hidden layer increases.
- SCG gives better performance with lesser number of neurons when the dataset is larger and for smaller datasets performs better with 4 neurons in the hidden dataset.

From the number of epochs taken to converge to the best validation performance we can draw the following conclusions:

- Levenberg in general converges faster and requires far lesser iterations even for larger datasets.
- Bayesian on the other hand takes more epochs to acheive the same MSE as Levenberg and is thus slower.
- SCG is a little better than Bayesian in that it takes fewer epochs than Bayesian and is almost on par with Levenberg in terms of the number of epochs it takes in general. But the drawback here is that the MSE attained here is much higher than that of Levenberg and it thus performs poorly.

6.3 Regression R-value

Regression R Values measure the correlation between outputs and targets. An R value of 1 means a close relationship, 0 a random relationship. [10]
Below are the Regression results obtained on training, test and overall on varying the number of neurons as well as the size of the dataset.

As can be seen from the tables 4, 5 and 6, the R-values are 0.98 for almost all the cases. Thus, our model exhibits a very strong correlation between predicted and actual values

Algorithm		R-value		
		2 neurons	4 neurons	8 neurons
Levenberg	2000 - 2015	0.98522	0.9853	0.98658
	2006 - 2015	0.98222	0.98252	0.98265
	2011 - 2015	0.98284	0.98362	0.98371
	2011	0.97868	0.97984	0.9812
Bayesian	2000 - 2015	0.98609	0.98673	0.98742
	2006 - 2015	0.9815	0.98309	0.98304
	2011 - 2015	0.9831	0.98461	0.98515
	2011	0.98102	0.98112	0.98356
SCG	2000 - 2015	0.98235	0.96597	0.96164
	2006 - 2015	0.95833	0.98277	0.98139
	2011 - 2015	0.95115	0.95757	0.97886
	2011	0.37673	0.96031	0.86483

Table 4: Training

Algorithm		R-value		
		2 neurons	4 neurons	8 neurons
Levenberg	2000 - 2015	0.98478	0.98446	0.98551
	2006 - 2015	0.98298	0.98275	0.98084
	2011 - 2015	0.98232	0.98238	0.98326
	2011	0.97963	0.98212	0.95344
Bayesian	2000 - 2015	0.98512	0.98631	0.98655
	2006 - 2015	0.98139	0.98065	0.9835
	2011 - 2015	0.98257	0.9815	0.98322
	2011	0.9631	0.97947	0.93502
SCG	2000 - 2015	0.98287	0.97059	0.96107
	2006 - 2015	0.95798	0.9817	0.98072
	2011 - 2015	0.95142	0.95858	0.97773
	2011	0.31291	0.96086	0.83047

Table 5: Test

Algorithm		R-value		
		2 neurons	4 neurons	8 neurons
Levenberg	2000 - 2015	0.9851	0.98508	0.98626
	2006 - 2015	0.98229	0.98243	0.98252
	2011 - 2015	0.98264	0.98351	0.9836
	2011	0.97844	0.97953	0.97743
Bayesian	2000 - 2015	0.98595	0.98666	0.9873
	2006 - 2015	0.98148	0.98274	0.98311
	2011 - 2015	0.98302	0.98413	0.98488
	2011	0.97853	0.98082	0.97669
SCG	2000 - 2015	0.98425	0.96697	0.96125
	2006 - 2015	0.95806	0.98252	0.98125
	2011 - 2015	0.95233	0.95744	0.97855
	2011	0.35509	0.95869	0.85808

Table 6: Overall

and all three models are performing well for our dataset. SCG exhibits poor performance only for small dataset with two neurons. This anomalous behaviour was observed even with MSE values.
On the whole we can conclude that all the three models are exhibiting a close relationship with R-values almost close to 1.

7 Real-world Insights

7.1 Key Takeaways from the project

In the present work an effort has been made to achieve a better understanding of the geomagnetic indices and their relationships with various plasma and field parameters during three different phases (rise, maximum and decay). The main conclusions can be drawn as follows:

- There is a strong correlation between Dst index and the solar wind parameters.
- Levenberg training algorithm gives the best performance and is the fastest running algorithm. Training in this algorithm stops automatically when generalization stops improving as indicated by an increase in the MSE of the validation samples. [11] [12]
- Bayesian Regularizaton algorithm typically takes more time, but can result in good generalization for difficult, small or noisy datasets. Training stops according to adaptive weight minimization (regularization). [13]
- Scaled Conjugate Gradient algorithm takes less memory. Training automatically stops when generalization stops improving, as indicated by an increase in the mean square error of the validation samples. [13]
- All the three algorithms have a high R-value.
- Thus, depending upon the performance needs, complexity of the data and the availability of resources we can choose one of the three algorithms.

7.2 Impact of our prediction in the real-world

The threat of space weather has become an increased concern for government officials around the world. Powerful solar flares and coronal mass ejections (CMEs) can devastate the world's interconnected power grids, airline operations, satellites and communications networks. [14] Different types of space weather can affect different technologies at Earth and the people who depend on these technologies. Solar flares can produce strong x-rays that degrade or block high-frequency radio waves used for radio communication during events known as Radio Blackout Storms. Solar Energetic Particles (energetic protons) can penetrate satellite electronics and cause electrical failure. These energetic particles also block radio communications at high latitudes in during Solar Radiation Storms [15]. Coronal Mass Ejections (CMEs) can cause Geomagnetic Storms at Earth and induce extra currents in the ground that can degrade power grid operations. Geomagnetic storms can also modify the signal from radio navigation systems (GPS and GNSS) causing degraded accuracy.

We have successfully been able to predict the Dst index which is an indicator of the presence of storms. With an R-value of nearly 0.98 in our model we can classify the presence of a storm with high confidence. Thus, necessary precautions can be taken and millions of dollars can be saved in terms of technology and communication failures.

8 Lessons Learned

- In this project we explored NARX model and learned about 3 different training algorithms that can be used in this model.
- We also learned about the use cases for these 3 training algorithms.
- Our understanding about solar storms and their impact on the real-world was further enhanced.
- We have explored state of the art implementations for predictions of solar storm.
- Most of the current implementations make use of ANN but none of them have used NARX for Dst index prediction. We thus chose NARX model in order to conduct an evaluation of how effective this model is in the prediction.

Things that could have been done differently:

- Given another chance, we would try predicting storms on the basis of other indices. For our project, we predicted Dst index values from the solar wind parameters, and for the predicted Dst value less than -100 nT, we declared it as a strom. Same could have been done for predicting other indices like Kp, AE, ap, ASY and SYM which could have given an even better estimate of predicting a storm.
- Also we could have done a comparative study to understand which input features (solar wind parameters) relate the best with each of the geomagnetic indices (Dst, Kp,AE etc)
- Instead of doing a binary classification of 0 (for Dst values greater than -100 nT i.e. no storm) and 1 (for Dst values less than -100 nT i.e. presence of storm), we could have made it a multi-class label problem. For extreme storms (peak Dst $<$ -200 nT), intense storms (-200 $<$ peak Dst $<$ 100 nT), moderate storms (100 nT $<$ peak Dst $<$ 50 nT) and weak storms (peak Dst $>$ 50 nT). This would have helped us in predicting the intensity of the storm.

References

[1] N. C. Joshi, N. S. Bankoti, S. Pande, B. Pande, and K. Pandey, "Behavior of plasma and field parameters and their relationship with geomagnetic indices during intense geomagnetic storms of solar cycle 23," *arXiv preprint arXiv:1003.2868*, 2010.

[2] W. Gonzalez, J. Joselyn, Y. Kamide, H. Kroehl, G. Rostoker, B. Tsurutani, and V. Vasyliunas, "What is a geomagnetic storm?," *Journal of Geophysical Research: Space Physics*, vol. 99, no. A4, pp. 5771–5792, 1994.

[3] T. N. Davis and M. Sugiura, "Auroral electrojet activity index ae and its universal time variations," *Journal of Geophysical Research*, vol. 71, no. 3, pp. 785–801, 1966.

[4] J.-G. Wu and H. Lundstedt, "Prediction of geomagnetic storms from solar wind data using elman recurrent neural networks," *Geophysical Research Letters*, vol. 23, no. 4, pp. 319–322, 1996.

[5] "Design time series narx feedback neural networks." https://www.mathworks.com/help/nnet/ug/design-time-series-narx-feedback-neural-networks.html.

[6] E. W. Weisstein, "Levenberg-marquardt method," 2000.

[7] F. Burden and D. Winkler, "Bayesian regularization of neural networks," *Artificial Neural Networks: Methods and Applications*, pp. 23–42, 2009.

[8] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural networks*, vol. 6, no. 4, pp. 525–533, 1993.

[9] J. Orozco and C. A. R. García, "Detecting pathologies from infant cry applying scaled conjugate gradient neural networks," in *European Symposium on Artificial Neural Networks, Bruges (Belgium)*, pp. 349–354, 2003.

[10] A. Tayarani, A. Baratian, M.-B. N. Sistani, M. R. Saberi, and Z. Tehranizadeh, "Artificial neural networks analysis used to evaluate the molecular interactions between selected drugs and human cyclooxygenase2 receptor," *Iranian journal of basic medical sciences*, vol. 16, no. 11, p. 1196, 2013.

[11] M. O. G. Nayeem, M. N. Wan, and M. K. Hasan, "Prediction of disease level using multilayer perceptron of artificial neural network for patient monitoring,"

[12] D. Liu, H. Zhang, M. Polycarpou, C. Alippi, and H. He, *Advances in Neural Networks–ISNN 2011: 8th International Symposium on Neural Networks, ISNN 2011, Guilin, China, May 29–June 1, 2011, Prodceedings*, vol. 3. Springer Science & Business Media, 2011.

[13] "Improve neural network generalization and avoid overfitting." https://www.mathworks.com/help/nnet/ug/improve-neural-network-generalization-and-avoid-overfitting.html.

[14] "Noaa takes 'huge step' forward in predicting solar storm impacts on earth." http://www.accuweather.com/en/weather-news/noaa-issues-regional-forecasts-for-solar-storms-for-first-time-with-new-geospacial-model/60788980.

[15] “Space weather impacts.” http://www.swpc.noaa.gov/impacts.