OBJECTIVE DISCRIMINATION OF GEOMAGNETIC DISTURBANCES AND PREDICTION OF DST INDEX BY ARTIFICIAL NEURAL NETWORKS*

S.A. Dolenko, I.N. Myagkova, V.R. Shiroky, I.G. Persiantsev
D.V. Skobeltsyn Institute of Nuclear Physics, M.V. Lomonosov Moscow State University, Moscow, Russia; e-mail: shiroky@srd.sinp.msu.ru

Abstract. Strong disturbances of the Earth’s magnetic field (geomagnetic storms) may have significant effect upon operation of engineering devices and well-being of people. Therefore, prediction of the state of magnetosphere is a very important problem. Geomagnetic disturbances (or magnetic storms) are one of important factors of space weather. In this study, we suggest an algorithm for objective discrimination of boundaries and different phases of magnetic storms based on the time series of hourly values of Dst-index. Two or three phases were marked for each storm: initial phase (optional), main phase, and recovery phase. With the help of the suggested algorithm, the boundaries of magnetic storms and their phases for the period from November 1997 till March 2014 have been marked automatically in an objective way. Then, neural network prediction of the value of Dst index by the parameters of solar wind and interplanetary magnetic field in L1 point and by preceding values of Dst index itself, has been performed. In this study, we perform detailed analysis of the results obtained for storm data, and suggest ways of improving existing approaches to neural network prediction of Dst-index.

INTRODUCTION

Space weather effects upon the near-Earth environment are due to dynamic changes in “Sun – solar wind – Earth” chain. The Earth’s magnetosphere is one of the key space environment domains affected by space weather. Developing successful prediction of the occurrence of geomagnetic disturbances (storms) is one of the important aims of solar-terrestrial physics research, since geomagnetic storms may have significant effect upon operation of engineering systems both in space and on Earth, and upon well-being of people.

The Dst index is one of the most important solar-terrestrial indices. It is used in numerous studies as a measure of intensity and temporal development of magnetic storms. The Dst index has been calculated at the World Data Center WDC-C2 at Kyoto, Japan (Geomagnetic Equatorial Dst index Home Page, http://wdc.kugi.kyoto-u.ac.jp/dstdir/index.html) since the International Geophysical Year, 1957, using data from four observatories at low to mid-latitudes; its hourly values are available online.

The main reason causing disturbances of the Earth’s magnetosphere are the fluxes of ionized particles from the Sun – the so-called solar wind (SW) (e.g. Akasofu and Chapman, 1972). For short-term prediction it is especially important to have operative information about the values of SW parameters. The input data usually used for Dst index prediction are the parameters of SW plasma and of interplanetary magnetic field (IMF), measured at Earth’s orbit. The data used in this work were obtained onboard ACE (Advanced Composition Explorer) spacecraft (http://www.srl.caltech.edu/ACE).

There are several different methods used to predict Dst index. For example, Space Research Institute (Russia) provides an advance warning about the future geomagnetic storm magnitude: real-time predictions of the geomagnetic storm magnitude are updated every hour and published at http://spaceweather.ru (Podladchikova and Petrukovich, 2012). The project of University of California, Berkeley (http://sprg.ssl.berkeley.edu/dst_index/welcome.html) produces a prediction of Dst index one hour ahead using data from spacecraft ACE, based on the modification of the empirical formula of Burton (Burton et al., 1975), the same way as it is done in the above-mentioned group at Space Research Institute. The Swedish Space Weather Center (http://src.irf.se/en/forecasts/) predicts the next hourly mean value of the Dst index (one hour forward in relation to the last entered data) using a recurrent Elman neural network. The project WINDMI Real-Time Dst and AL indices at the CCMC (http://ccmc.gsfc.nasa.gov/cgi-bin/WINDMIPred.cgi) provides a Dst prediction, which is performed using a physical model based on the calculation of ring currents in the magnetosphere-ionosphere system, and which also receives as input the data from the spacecraft ACE (Patra et al., 2011). A modern model used to predict the Dst index 1 hour ahead proposed in (Revallo et

* This study was supported by the Russian Foundation for Basic Research (RFBR), project no. 14-01-00293.
OBJECTIVE DISCRIMINATION OF GEOMAGNETIC DISTURBANCES

Correct prediction of Dst index is most valuable during geomagnetic disturbances (GMD), especially during strong magnetic storms. Since such events are relatively rare, it is necessary to have an opportunity to extract data sets with samples corresponding only to the disturbed state of magnetosphere. Such data sets may be used for adequate assessment of prediction quality, and also in ANN training – to make the ANN perform better during GMD.

A geomagnetic disturbance consists of three phases. The main phase is characterized by a rapid fall of Dst index from background values (-15…-20 nT to 3…5 nT) down to large negative values (lower than -30 nT for weak GMD, -50 nT for moderate GMD, -100 for strong GMD etc.). The main phase is followed by the recovery phase, during which the value of the Dst index gradually returns to the background level; the duration of this phase may reach several days. The main phase is preceded by the initial phase, which often starts with the so-called sudden storm commencement (SSC). We suppose, that in this case during the initial phase the Dst index takes on positive values greater than 3…5 nT, and here we shall consider that detecting such positive values corresponds to detecting initial phase with SSC.

While the point of the minimum value of Dst index is considered to separate the main phase from the recovery phase, there are no commonly accepted criteria to determine the exact hour of GMD start, GMD end, or the border between the initial and the main phases. To mark such points, a human expert usually analyses not only the time series (TS) of the Dst index, but also the behavior of SW and IMF parameters. Therefore, discrimination of GMD requires much human effort, and it is subjective, as different experts may come to different conclusions.

So there is a demand for an algorithm for objective discrimination of GMD and their phases, which would require no human expertise. Here we report for the first time the results of such an algorithm created by the authors. Note that this version of the algorithm analyses only the TS of the Dst index itself and it does not analyze the behavior of SW and IMF parameters, thus having natural limits in its efficiency. However, the algorithm uses several passes over the data, and its parameters were developed and fine-tuned in such a way that our human expert agrees with most of the results of GMD marking by the algorithm.

This version of the algorithm has seven parameters:

- $A_{min}$ is the maximum negative value of Dst index in nT required for the GMD to be detected (a negative integer) – minimal amplitude of detected GMD.
- $t_+$ is the upper boundary of background Dst values in nT (an integer).
- $t_-$ is the lower boundary of background Dst values in nT (an integer).
- $s$ is the TS smoothing parameter in hours (a non-negative integer). If $s=0$, smoothing is turned off.
- $B_{short}$ is the short merging parameter in hours (a non-negative integer).
- $B_{long}$ is the long merging parameter in hours (a non-negative integer).
- maxRewind is the starting point search parameter in hours (a non-negative integer).

The main stages of the algorithm are the following.

1) Time series smoothing. A working TS is created, each point of which assumes a value equal to the average value of the initial Dst TS over a window $2s+1$ points wide centered at this point in time.

2) Preliminary marking. Each point of the working TS with amplitude $d$ is marked as belonging to one of the four types:

- Type A ($d \geq t'$) – possible initial phase point;
- Type B ($t_+ > d > t_-$) – possible background point;
- Type C ($t_+ \geq d > A_{min}$) – possible main phase or relaxation phase point;
- Type D ($d \leq A_{min}$) – definite main phase or relaxation phase point.

3) Interval creation. All adjacent points of the same type (except type B) are joined into intervals of this type.
4) Interval joining. Intervals separated with number of type B points less or equal to $B_{short}$ (or $B_{long}$, depending on types of these intervals), are merged into joint intervals.

5) Event determination. Each of the joint intervals containing at least one type D point is considered to be a separate event (GMD). All the other joint intervals are disregarded.

6) Initial phase refinement. If a joint interval includes at least one type A point, the corresponding GMD is considered to have the initial phase with SSC. A special iterative search procedure to the left of the current beginning of the joint interval with the maxRewind parameter is used to locate the first point of GMD as a local minimum or a first positive value point in the working TS.

7) Starting point refinement. For joint intervals considered having no initial phase with SSC, a special iterative search procedure to the left of the current beginning of the joint interval with the maxRewind parameter is used to locate the first point of GMD as a local maximum or a first negative value point in the working TS. If type A points are found during this search, then the GMD is considered to have the initial phase with SSC, and the algorithm goes back to Stage 6.

8) Output of the results. Each of the finally obtained joint intervals is treated as a separate GMD. The algorithm outputs the list of found GMDs, specifying for each GMD: minimum Dst index value, hours of GMD start, of Dst minimum, of GMD end, and presence of the initial phase with SSC. In a separate log file, each point is marked with the ID of GMD it belongs to and with the GMD phase it belongs to.

A detailed description of the algorithm and the results of its work will be published separately.

The optimal parameter values were fine-tuned to be the following: $t_+ = 5$ nT, $t_- = -20$ nT, $s = 1$ hour, $B_{short} = 2$ hours, $B_{long} = 4$ hours, maxRewind = 24 hours. For this study, $A_{min}$ was set to -30 nT (weak and stronger GMD), but this value can be changed if it is necessary to mark only strong or only extreme GMD.

For the total period from October 22, 1997 till March 31, 2014, the algorithm has marked 867 GMDs, 300 of them having the initial phase with SSC.

As an example, the Table displays the list of GMD marked by the algorithm for the period from 0-1h UT March 1, 2012 till 23-24h UT April 30, 2012. To avoid confusion, each hour is specified by both its beginning and its end. The time in the Table is UT. Fig. 1 displays the Dst index TS for the same period.

<table>
<thead>
<tr>
<th>№</th>
<th>Min. of Dst, nT</th>
<th>Start of GMD at: yyyy-mm-dd_hh-b</th>
<th>Dst minimum at: yyyy-mm-dd_hh-b</th>
<th>End of GMD at: yyyy-mm-dd_hh-b</th>
<th>Init. phase with SSC?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-34</td>
<td>2012-02-29_23-24</td>
<td>2012-03-02_01-02</td>
<td>2012-03-02_05-06</td>
<td>yes</td>
</tr>
<tr>
<td>2</td>
<td>-32</td>
<td>2012-03-03_18-19</td>
<td>2012-03-04_02-03</td>
<td>2012-03-04_05-06</td>
<td>no</td>
</tr>
<tr>
<td>3</td>
<td>-74</td>
<td>2012-03-07_02-03</td>
<td>2012-03-07_09-10</td>
<td>2012-03-08_08-09</td>
<td>no</td>
</tr>
<tr>
<td>4</td>
<td>-131</td>
<td>2012-03-08_10-11</td>
<td>2012-03-09_08-09</td>
<td>2012-03-11_11-12</td>
<td>yes</td>
</tr>
<tr>
<td>5</td>
<td>-50</td>
<td>2012-03-12_09-10</td>
<td>2012-03-12_16-17</td>
<td>2012-03-14_09-10</td>
<td>yes</td>
</tr>
<tr>
<td>6</td>
<td>-33</td>
<td>2012-03-14_11-12</td>
<td>2012-03-15_02-03</td>
<td>2012-03-15_07-08</td>
<td>no</td>
</tr>
<tr>
<td>7</td>
<td>-74</td>
<td>2012-03-15_13-14</td>
<td>2012-03-15_20-21</td>
<td>2012-03-17_11-12</td>
<td>no</td>
</tr>
<tr>
<td>8</td>
<td>-43</td>
<td>2012-03-17_13-14</td>
<td>2012-03-18_02-03</td>
<td>2012-03-18_23-24</td>
<td>no</td>
</tr>
<tr>
<td>9</td>
<td>-55</td>
<td>2012-03-27_06-07</td>
<td>2012-03-28_04-05</td>
<td>2012-03-28_19-20</td>
<td>yes</td>
</tr>
<tr>
<td>10</td>
<td>-30</td>
<td>2012-04-02_14-15</td>
<td>2012-04-02_19-20</td>
<td>2012-04-03_03-04</td>
<td>no</td>
</tr>
<tr>
<td>11</td>
<td>-54</td>
<td>2012-04-04_21-22</td>
<td>2012-04-05_07-08</td>
<td>2012-04-05_17-18</td>
<td>no</td>
</tr>
<tr>
<td>12</td>
<td>-31</td>
<td>2012-04-07_02-03</td>
<td>2012-04-07_10-11</td>
<td>2012-04-07_12-13</td>
<td>no</td>
</tr>
<tr>
<td>14</td>
<td>-108</td>
<td>2012-04-23_02-03</td>
<td>2012-04-24_04-05</td>
<td>2012-04-26_04-05</td>
<td>yes</td>
</tr>
</tbody>
</table>

Fig. 1. Dst index time series from March 1 to April 30, 2012.
INPUT DATA FOR PREDICTION

As input data, TS of hourly values of the following physical quantities were used:

1) SW parameters in Lagrange point L1 between the Earth and the Sun:
   • Speed \( v \) (in km/s); protons density \( n_p \) (in cm\(^{-3}\)); temperature \( T_{rr} \) (in K).

   The data used was from ACE spacecraft, measured by SWEPAM device.

2) IMF vector parameters in Lagrange point L1 (in nT) in GSM system:
   • \( B_x \), \( B_y \), \( B_z \) (IMF x-, y-, and z-components), \( B \) magnitude (IMF modulus).

   The data used here was also from ACE spacecraft, MAG device.

3) Geomagnetic indexes:
   • Equatorial geomagnetic index Dst (measured in nT).

   The data used was from World Data Centre for Geomagnetism in Kyoto.

4) Earth movement phase parameters (two phases of harmonic function for each movement):
   • With the period equal to that of the Earth’s revolution round the Sun;
   • With the period equal to that of the Earth’s rotation around its axis.

   The considered data period was from November 1997 to the end of March 2014.

   Due to specifics of data obtained from measurements onboard spacecraft, there were some missing values in SW and IMF parameters. If the duration of the gap did not exceed 10 hours, the missing values were replaced by linear interpolation between the last known data value before the gap and the first known data value after the gap, separately for each TS.

   After interpolation, delay embedding of all TS (except Earth movement phase parameters) for 24 hour depth was used, to account for the previous history of input features. Thus, each data pattern was a point in 204-dimensional feature space (8 TS \( \times \) 25 hours + 4 TS), assigned to the latest point of the 25-hour window.

NEURAL NETWORK MODELS

In this study, we used perceptron type ANN with one hidden layer, 204 neurons in the input layer, 32 neurons in the hidden layer, and 1 output neuron. For each model, 5 ANN were trained that were identical except for various initialization of weights. The answers of these 5 ANN were averaged, and the obtained value was considered to be the output value of the model. Prediction horizon for all models was one hour.

All the available samples were divided into training, test and examination data sets. The training set was used to determine the error for weights adjustment, the test set was used for recurrent testing during network training to provide timely termination of the training (to prevent overtraining), and the examination set was used to test the trained network on out-of-sample data to assess the network quality. The data from 1997 to 2009 (105187 samples), which approximately correspond to one full solar activity cycle, were randomly divided into training and test sets with 3:1 ratio respectively. All the remaining samples (from 2010 to 2014) were used as the examination set (37178 samples).

Most interesting is the quality of prediction for the samples corresponding to the disturbed state of magnetosphere. So, a separate “disturbed” data set has been extracted, which included all the samples marked as belonging to GMD by the above-described algorithm. Intersection of the disturbed set with training, test, and examination data sets gave the three more special disturbed sets that provided the possibility to assess network quality for GMD samples only.

The statistical indicators used to assess model quality on various data sets were root mean squared error (RMSE) and coefficient of multiple determination \( R^2 \).

The following five models were compared.

1) **Full model** trained on full data sets, including samples corresponding to as disturbed, as non-disturbed state of the Earth’s magnetosphere. Delay embedding depth was 24 hours as described above (204 input features), the number of samples in the training set for this model was 78890.

2) **Disturbed model** trained using the special disturbed data sets as described above. Delay embedding was also to the depth of 24 hours (204 input features). The number of samples in the training set was 24157.

3) **Full model** completely similar to the first one except for delay embedding depth that was set to 8 hours. So, the number of input features was 8 TS \( \times \) 9 hours + 4 TS = 76, the number of training samples was 78890. Note that maximum of correlation between Dst value and delayed values of some input time series is achieved at delays greater than 8 hours, so this means some reduction in the amount of information.

4) **Intermediate model** completely similar to the first one except for the fact that it had as many samples in all the sets as the second model; however, these samples were randomly selected from full data sets,
therefore they contained both GMD and non-GMD samples. So this model also had 204 input features and 24157 training samples, just like the second one.

5) **Trivial model** used as a reference point in model assessment. For the trivial model, the predicted value of the Dst index is set to its current value. Therefore, it is completely useless, but its statistical indicators are often quite good, so comparison with this model is necessary for real estimation of the prediction quality of any more sophisticated model.

**RESULTS AND DISCUSSION**

In this study, we compare the performance of the five described models on the special disturbed datasets. Fig. 2 compares values of RMSE in nT, and Fig. 3 compares values of $R^2$.

![Fig. 2. RMSE of the models on the special disturbed datasets.](image)

![Fig. 3. $R^2$ of the models on the special disturbed datasets.](image)

From Fig. 2, 3, one can draw the following conclusions.

1) The results on training and test sets are in general alike, for all models except the disturbed one. This means that for the non-GMD data that was used to train these models, the test set was representative enough, and training termination at minimum error on test set succeeded in preventing ANN overtraining. This is not the case with the disturbed model, which gets somewhat overtrained; this may mean that the disturbed data set is not large enough to provide due representativity even for random choice of the test set.

2) The examination set displays worse results for all the models including the trivial one. This may mean that the present solar cycle starting from 2010 is somewhat different from the preceding one – at least, the storms are much more volatile.

3) Comparing full model (24 h) with intermediate model, one can see that in spite of the relatively large number of training samples (24157) even in respect to the number of input features, they are not
enough to fully describe the extremely complex object that is studied. 3+ fold decrease in the number of samples (even with random selection of these samples) the ANN is trained on results in noticeable degradation of its performance, in particular, on the disturbed data sets.

4) Comparison of the intermediate and the disturbed models, both of which were trained on equal number of samples, shows that the model trained on GMD data performs on disturbed data sets better, than the model trained on all kinds of data. This is not surprising – more interesting is small difference of these two models on test and examination data sets. This may mean that the non-GMD data bears a significant amount of information about the dynamics of the magnetosphere in the whole, and this information can be acquired by the ANN.

5) Comparison of the two versions of the full model with different time series embedding depth shows that the reduction in network performance for smaller embedding depth is insignificant. However, this does not necessarily mean that 8 hours are enough. In fact, what we see is the result of two concurring processes – a) ANN performance degradation due to lack of information brought by the features corresponding to time delays larger than 8 hours, and b) improvement of ANN performance due to a significant reduction in the number of input features. Proper selection of the most significant input features from the whole range of delays from 0 to 24 hours may result in a configuration having as little input features as the 8-hour model, but performing better. Elaboration of an algorithm of such selection should be the aim of future studies.

It should be also noted that all the considered models much outperform the trivial model.

CONCLUSION

An algorithm for objective discrimination of boundaries and different phases of geomagnetic disturbances, based on the time series of hourly values of Dst geomagnetic index, has been elaborated. With the help of the suggested algorithm, the boundaries of GMD and their phases for the period from November 1997 till March 2014 have been marked automatically in an objective way.

A neural network prediction of the value of Dst index by the parameters of solar wind and interplanetary magnetic field in L1 point and by preceding values of Dst index itself, has been performed. The networks were trained on data of the complete cycle of solar activity from 1997 till 2009, and examined on samples from 2010 till March 2014 belonging to GMD, marked by the above algorithm.

It has been demonstrated that the best results on the GMD samples are provided by the ANN model trained on the whole data array disregarding presence or absence of GMD. The situation may change with time, when the representativity of the disturbed set will increase together with total number of samples, as an ANN trained only on GMD samples outperforms an ANN trained on the same amount of general data.

The prediction may be also improved with proper selection of significant input features, as even a simple reduction in delay embedding depth equilibrates decrease in the amount of information with the decrease of the number of input features.

ACKNOWLEDGEMENT

The authors thank the teams of ACE project and WDCC2 Kyoto for the possibility to access and use the data of their experiments.

REFERENCES