Statistically predicting Dst without satellite data

A. S. Parnowski

Space Research Institute NASU & NSAU, 40 prosp. Akad. Glushkova, Kyiv-187, 03680 MSP, Ukraine

In this paper we construct a regression relationship for predicting Dst 1 hour ahead. Our model uses only previous Dst values. This regression is totally unbiased and does not rely on any physical model, except for the fact that Dst somehow contains the information on the recurrent geomagnetic storms. This regression has the prediction efficiency of 0.964, linear correlation with official Dst index of 0.982, and RMS of 4.52 nT. These characteristics are inferior only to our other model, which uses satellite data and provides the prediction efficiency of 0.975, linear correlation with official Dst index of 0.986, and RMS of 3.76 nT. This makes it quite suitable for prediction purposes when satellite data are not available.
Key words: Space weather, statistical model, Dst prediction.

1. Introduction

Space weather prediction is one of the main tasks of modern space research. The necessity of such activities was well understood for a long time (Marubashi, 1989). Space weather prediction activities divide into two large categories: prediction of space weather directly in space, and prediction of space weather manifestations on the Earth. The first category is mostly important for planning of space missions, predicting and evading hardware failures of spacecraft due to arcing in electronic components, and assuring astronaut safety with respect to radiation hazard. These tasks mainly require prediction of energetic particle fluxes. The second category deals with influence of space weather on power grid operation, radio communications, and health of people, especially those with cardiovascular diseases. These tasks mainly require prediction of geomagnetic disturbances. This article will focus on space weather prediction on the Earth.

There are many quantitative indices of geomagnetic activity. The most widely used of them are storm-time disturbance Dst and planetary geomagnetic activity index Kp. Dst is more convenient for prediction purposes, because it directly equals the disturbance of H-component on the Earth measured in gammas ($1\gamma = 10^{-4}\,\mathrm{Gs} = 1\,\mathrm{nT}$). It is averaged over several low- and mid-latitude magnetometer stations and is usually associated with the westward ring current, which appears during the storm at 4–$8R_\mathrm{E}$, although this association was strongly criticized by e.g. Campbell (1996). At the same time, Kp is an integral and more artificial characteristic of the overall level of geomagnetic disturbance. In this paper Dst is used for space weather prediction.

Space weather prediction is a challenging and nontrivial activity (Joselyn, 1995; Li *et al.*, 2003). Since Burton *et*

al. (1975) published their pioneering work, many authors tried to forecast space weather indices. Papers (Kugblenu *et al.*, 1999; Watanabe *et al.*, 2002; Wing *et al.*, 2005; Pallocchia *et al.*, 2006) featured neural network approach; papers (Zhou and Wei, 1998; Balikhin *et al.*, 2001; Harrison and Drezet, 2001) incorporated adaptive filtering; papers (Rangarajan and Barreto, 1999; Oh and Yi, 2004; Wei *et al.*, 2004; Johnson and Wing, 2004) applied statistical methods; papers (Burton *et al.*, 1975; Valdivia *et al.*, 1996; O'Brien and McPherron, 2000a, b; Temerin and Li, 2002; Ballatore and Gonzales, 2003; Cid *et al.*, 2005; Siscoe *et al.*, 2005; Temerin and Li, 2006) used empirical models; and papers (Dryer *et al.*, 1984; Raeder *et al.*, 2001) developed global MHD simulations.

The best results for 1-hour prediction were achieved by Temerin and Li (2002, 2006), who used an empirical model. They achieved the prediction efficiency of 0.91, linear correlation of 0.95 and RMS of 6.4 nT. Neural network approach provides short-term predictions up to 4 hours in the paper (Wing *et al.*, 2005). It experiences significant difficulties predicting geomagnetic storms with Kp > 5, though. Adaptive filtering seems more successful being able to provide 8-hour predictions in the paper (Harrison and Drezet, 2001). However, in the papers, which incorporate adaptive filtering, the volume of the dataset usually does not exceed 6 months of data (4380 points), which is not enough to correctly describe long-time variations in geomagnetic activity, caused, e.g., by 11-year solar cycle. Statistical methods give interesting results, but were rarely used for prediction, and much more often for developing and constraining empirical models (Johnson and Wing, 2004). Empirical models were the most often used, and provided some of the best 1-hour predictions. Most of them are improvements of the empirical relationship proposed in a pioneering paper by Burton *et al.* (1975), who analysed the ring current injection and decay. However, their model suffered from the lack of solar wind data and poor physical understanding of solar wind-magnetosphere interaction at that time. Global MHD simu-
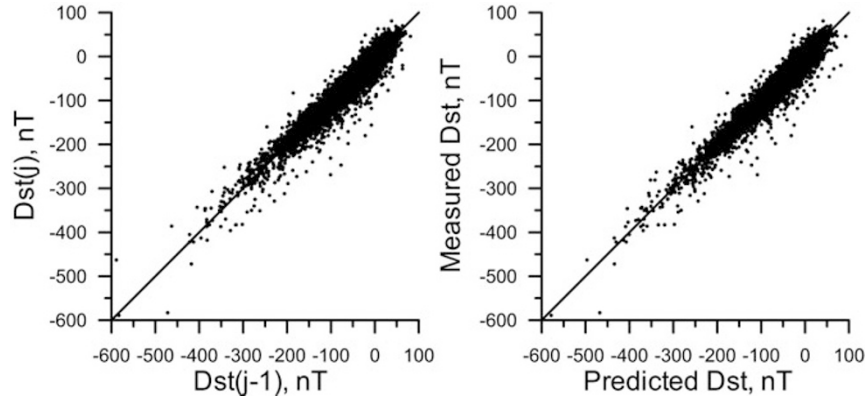
Fig. 1. Scatter plots of measured Dst versus Dst 1 hour ago (on the left) and versus predicted Dst (on the right).

lations give the longest prediction times but fail to correctly describe kinetics in boundary layers and ballooning instabilities, which are believed to be responsible for the substorm onset.

Here we use the same method as in our other article (Parnowski, 2008), which combines statistical and empirical approaches. We payed attention mostly to 1-hour prediction, though we obtained a 9-hour prediction as well. We predicted Dst 1 hour ahead because the temporal resolution of the dataset was 1 hour, so we just predicted the next value in the series. Besides, longer prediction times resulted in predicted value being shifted in time. For prediction we use only that information, which is available at the moment when prediction is made, i.e. 1 hour prior to the predicted value. We will reference to this value as "1 hour ago". We determine the quality of prediction by 3 values: residual mean square (RMS), prediction efficiency (PE), defined as $[1 - (\text{mean squared residual})/(\text{variance of data})]$ (Temerin and Li, 2002), and the linear correlation coefficient (LC) between the prediction and Dst. In the article (Parnowski, 2008) we constructed a regression relationship, which provides PE = 0.975, LC = 0.986, and RMS = 3.76 nT.

However, this relationship requires satellite data to be continuous for the previous 20 hours. Thus, when the data contain a gap for some reason, we are unable to predict Dst for the next 20 hours. For this reason, we need an 'emergency' regression, which would operate without satellite data. Besides, satellite data are often missing during very strong geomagnetic storms, which are the most interesting events. Such a regression will be constructed in this article.

2. Data, Routine and Results

We used the OMNI 2 database, available at NSSDC (http://nssdc.gsfc.nasa.gov/omniweb/). It contains IMF, solar wind and geomagnetic data, averaged over 1-hour intervals (49 parameters in total, starting from Jan 1, 1963). The complete 43-year Dst time series given therein is continuous and features an eye-visible 27-day and 11-year periodicity, which hints for strong dependence on solar activity.

We seek Dst in a regression form

$$\text{Dst}(j) = \sum_i C_i x_i,$$

where j is the current step (number of hours since Jan 1,

1963), C_i are the regression coefficients, and x_i are the regressors, which are functions of input quantities and their combinations. Values of C_i are determined by the least square method with equal statistical weights of all points, and the statistical significance of the regressors–by Fisher test (Fisher, 1954; Hudson, 1964).

The initial number of regressors was deliberately excessive to let Fisher test select the most statistically significant of them. This was done in the following way. After processing the data with the least square method, Fisher significance parameter F was determined for each regressor. All F values were compared to the values 2.7055, 3.84, 5.02, 6.635, 7.879, 10.83 and 12.1, which correspond to statistical significance of 90, 95, 97.5, 99, 99.5, 99.9 and 99.95% respectively. Then, insignificant regressors were rejected and the routine was repeated until all the regressors were significant. We chose the minimal significance level of 90%. In contrast to empirical models we do not add fitting parameters and all the regressors have obvious physical meaning. The described routine was applied to the complete 43-year dataset sans rejects. More details on the routine can be found in the paper (Parnowski, 2008).

First, we determined which previous Dst values are statistically significant. For this purpose, we constructed a regression

$$\text{Dst}(j) = C_0 + \sum_{i=1}^{N} C_i \, \text{Dst}(j - i),$$

where N is the oldest Dst value; we reached the value $N = 900$. We found that there are statistically significant values as far as 801 hours ago (33 days and 9 hours). The statistical significance of this oldest value is over 99.9%. However, it is possible that there are even older statistically significant values. A similar situation was reported by Johnson and Wing (2004) regarding Kp: "the significance is often quite large for extended periods of time (10–20 days)". This might be related in some way to recurrent geomagnetic storms, but some additional research is required before final explanation could be given to this phenomenon. This will be done in a future article.

After determining which previous Dst values are statistically significant, we added nonlinear terms. We tried different powers of the most significant terms and their products.
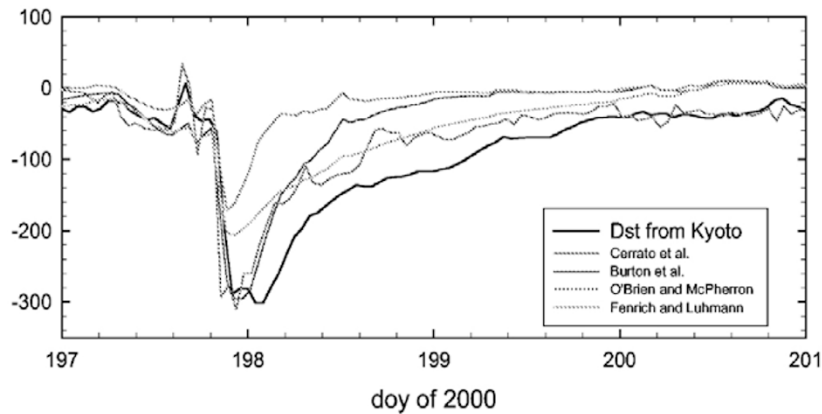
Fig. 2. Experimental Dst index data and theoretical predictions from several models. Reprinted from (Cid *et al*., 2005). "Dst from Kyoto" stands for the official Dst index from WDC-B.
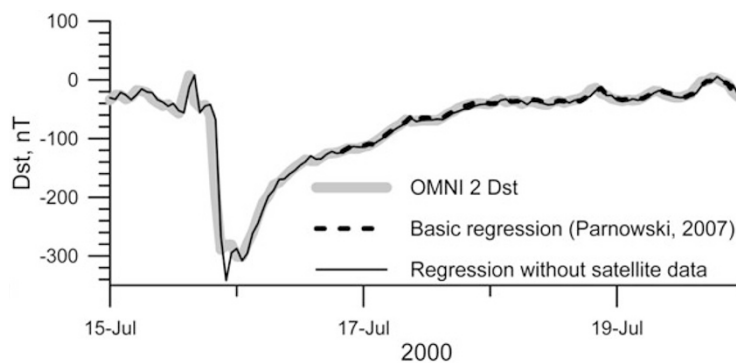

Fig. 3. Our prediction for the same period of time as on Fig. 2. Satellite data is missing in the left part of the plot.

Thus, we constructed a regression, which consisted only of Dst terms and a constant regressor. Its characteristics are: $PE = 0.964$, $LC = 0.982$, $RMS = 4.53$ nT.

To illustrate predictive capabilities of our model we present several figures: Fig. 1 shows scatter plots of measured Dst versus Dst 1 hour ago (it is the simplest possible prediction model) on the left and versus prediction on the right; Fig. 2 shows predictions by Cerrato *et al*. (2004), Fenrich and Luhmann (1998), O'Brien and McPherron (2000b), and Burton *et al*. (1975) for Jul 15–19, 2000; Fig. 3 shows our prediction for the same period of time.

More comparison with other models can be found in the paper (Parnowski, 2008).

3. Conclusion

It appeared possible to predict Dst 1 hour ahead using only its previous values. This hints for recurrent behaviour of geomagnetic activity. In terms of prediction efficiency and linear correlation with the official Dst index this model is inferior to the model, which uses satellite solar wind data. However, satellite data is often missing during strong geomagnetic storms and this model can be used to fill the gap in predicted Dst time series. Thus, a combination of a regression model using satellite data with the model obtained in this article can provide accurate on-line operational Dst forecast.

Of course, larger prediction times are desirable, but 1-hour prediction is still useful. For example, magnetometers can be switched to high temporal resolution mode, sensitive equipment can be turned off, etc.

4. Summary

In this article we obtained the following results:

1. We derived a regression, which relies only on previous Dst values.
2. It allows predicting Dst 1 hour ahead with $PE = 0.964$, $LC = 0.982$ and $RMS = 4.53$ nT. Thus, it is very convenient for on-line Dst prediction when satellite data are not available.
3. Previous Dst values are statistically significant up to 801 hours ago and possibly more.
4. Nonlinear terms appeared to be very significant.

References

Balikhin, M. A., O. M. Boaghe, S. A. Billings, and H. St. C. K. Alleyne, Terrestrial magnetosphere as a nonlinear resonator, *Geophys. Res. Lett.*, **28**, 1123–1126, 2001.

Ballatore, P. and W. D. Gonzalez, On the estimates of the ring current injection and decay, *Earth Planets Space*, **55**, 427–435, 2003.

Burton, R. K., R. L. McPherron, and C. T. Russell, An empirical relationship between interplanetary conditions and Dst, *J. Geophys. Res.*, **80**, 4204–4214, 1975.

Campbell, W. H., Geomagnetic storms, the Dst ring-current myth and lognormal distributions, *J. Atm. Terr. Phys.*, **58**, 1171–1187, 1996.

Cerrato, Y., E. Saiz, C. Cid, and M. A. Hidalgo, Geomagnetic storms: their sources and a model to forecast Dst index, in *Lecture notes and essays in Astrophysics*, pp. 131–142, 2004.

Cid, C., E. Saiz, and Y. Cerrato, Physical models to forecast the Dst index: comparison of results, *Proc. Solar Wind 11—SOHO 16 "Connecting Sun and Heliosphere", Whistler, Canada 12–17 June 2005 (ESA SP-592, September 2005)*, 116–119, 2005.

Dryer, M., S. T. Wu, G. Gislason, S. M. Han, Z. K. Smith, J. F. Wang, D. F. Smart, and M. A. Shea, Magnethydrodynamic modelling of interplanetary disturbances between the Sun and Earth, *Astrophys. Space Sci.*, **105**, 187–208, 1984.

Fenrich, R. R. and J. G. Luhmann, Geomagnetic response to magnetic clouds of different polarity, *Geophys. Res. Lett.*, **25**, 2999, 1998.

Fisher, R. A., *Statistical methods for research workers, Twelfth edition*, London, Oliver and Boyd, 1954.

Harrison, R. F. and P. M. Drezet, The application of an adaptive nonlinear systems identification technique to the on-line forecast of Dst index, *Proc. Les Woolliscroft memorial Conf. / Sheffield Space Plasma Meeting: Multipoint measurements versus theory, Sheffield, UK, Apr 24–26, 2001 (ESA SP-492)*, 141–146, 2001.

Hudson, D. J., *Statistics Lectures on Elementary Statistics and Probability*, Geneva, CERN, 1964.

Johnson, J. R. and S. Wing, A cumulant-based analysis of nonlinear magnetospheric dynamics, *Report PPPL-3919rev*, http://www.pppl.gov/pub_report/2004/PPPL-3919rev.pdf, 2004.

Joselyn, J. A., Geomagnetic activity forecasting—the state-of-the-art, *Rev. Geophys.*, **33**, 383–401, 1995.

Kugblenu, S., S. Taguchi, and T. Okuzawa, Prediction of the geomagnetic storm associated Dst index using an artificial neural network algorithm, *Earth Planets Space*, **51**, 307–313, 1999.

Li, X., M. Temerin, D. N. Baker, G. D. Reeves, D. Larson, and S. G. Kanekal, The predictability of the magnetosphere and space weather, *EOS*, **84**, 2003.

Marubashi, K., The space weather forecast program, *Space Sci. Rev.*, **51**, 197–214, 1989.

O'Brien, T. P. and R. L. McPherron, Forecasting the ring current index Dst in real time, *J. Atm. Sol.-Terr. Phys.*, **62**, 1295–1299, 2000a.

O'Brien, T. P. and R. L. McPherron, An empirical phase-space analysis of ring current dynamics: Solar wind control of injection and decay, *J. Geophys. Res.*, **105**, 7707, 2000b.

Oh, S. Y. and Y. Yi, Relationships of the solar wind parameters with the magnetic storm magnitude and their association with the interplanetary shock, *J. Korean Astron. Soc.*, **37**, 151–157, 2004.

Pallocchia, G., E. Amata, G. Consolini, M. F. Marcucci, and I. Bertello, ANN prediction of the Dst index, *Mem. S.A.It. Suppl.*, **9**, 120–122, 2006.

Parnowski, A. S., Dst prediction using the linear regression analysis, *Kosmichna Nauka i Technologiya*, 2008 (accepted, in Russian).

Raeder, J. *et al.*, Global simulation of the Geospace Environment Modeling substorm challenge event, *J. Geophys. Res.*, **106**, 381–396, 2001.

Rangarajan, G. K. and L. M. Barreto, Use of Kp index of geomagnetic activity in the forecast of solar activity, *Earth Planets Space*, **51**, 363–372, 1999.

Siscoe, G., R. L. McPherron, M. W. Liemohn, A. J. Ridley, and G. Lu, Reconciling prediction algorithms for Dst, *J. Geophys. Res.*, **110**, A02215, doi:10.1029/2004JA010465, 2005.

Temerin, M. and X. Li, A new model for the prediction of Dst on the basis of the solar wind, *J. Geophys. Res.*, **107**, 1472, doi:10.1029/2001JA007532, 2002.

Temerin, M. and X. Li, Dst model for 1995–2002, *J. Geophys. Res.*, **111**, A04221, doi:10.1029/2005JA011257, 2006.

Valdivia, J. A., A. S. Sharma, and K. Papadopoulos, Prediction of magnetic storms by nonlinear models, *Geophys. Res. Lett.*, **23**, 2899–2902, 1996.

Watanabe, S., E. Sagawa, K. Ohtaka, and H. Shimazu, Prediction of the Dst index from solar wind parameters by a neural network method, *Earth Planets Space*, **54**, 1263–1275, 2002.

Wei, H. L., S. A. Billings, and M. A. Balikhin, Analysis of the geomagnetic activity of the Dst index and self-affine fractals using wavelet transforms, *Nonlinear Process. Geophys.*, **11**, 303–312, 2004.

Wing, S., J. R. Johnson, J. Jen, C.-I. Meng, D. G. Sibeck, K. Bechtold, J. Freeman, K. Costello, M. Balikhin, and K. Takahashi, Kp forecast models, *J. Geophys. Res.*, **110**, A04203, doi:10.1029/2004JA010500, 2005.

Zhou, X.-Y. and F.-S. Wei, Prediction of recurrent geomagnetic disturbances by using adaptive filtering, *Earth Planets Space*, **50**, 839–845, 1998.

A. S. Parnowski (e-mail: dyx@ikd.kiev.ua)