

STATISTICAL APPROACH TO Dst PREDICTION

A. S. Parnowski

Space Research Institute NASU & NSAU,

40, prosp. Akad. Glushkova, Kyiv-187, 03680 MSP, Ukraine

(Received January 14, 2008; received in final form April 10, 2008)

Using an unbiased statistical approach we construct a phenomenological model of solar wind — magnetosphere interaction. This model allows measuring relative geoeffectiveness of different physical parameters, uncovering some empirical dependencies, accurately predicting the geomagnetic activity up to 3 hours ahead, and estimating the geomagnetic activity up to 9 hours ahead. For 1-hour forecast we have linear correlation $LC = 0.987$ and prediction efficiency $PE = 0.975$. For 3-hour forecast we have $LC = 0.906$ and $PE = 0.899$. For 9-hour forecast we have $LC = 0.820$ and $PE = 0.711$.

Key words: space weather, linear regression, geoeffective parameters, solar-terrestrial physics, solar wind — magnetosphere interaction, magnetic storms.

PACS number(s): 02.70.Rr, 94.05.Sd, 94.05.Sx, 94.30.Lr, 94.30.Vf

I. 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 [1]. 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 the influence of space weather on power grid operation, radio communications, etc. 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 the 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 or nanoteslas (1 gamma = 1 nT). It is averaged over 4 low- and mid-latitude magnetometer stations and is usually associated with the westward ring current, which appears during the storm at 4-8 RE, although this association was strongly criticized, e. g. by Campbell [2]. At the same time, Kp is an integral and more artificial characteristic of the overall level of geomagnetic disturbance. Dst was used for space weather prediction earlier by many authors [3–17]. However, Kp index was also used for space weather prediction in a number of papers [18–20]. In this paper Dst is used for space weather prediction.

Space weather prediction is a challenging and non-trivial activity [21]. For this reason different approaches were used. Papers [5, 12, 17, 20] featured neural network approach; papers [8, 9, 18] incorporated adaptive filter-

ing; papers [14, 19, 22, 23] applied statistical methods; papers [3, 4, 6, 7, 10, 11, 13, 15, 16] used empirical models; and papers [24, 25] developed global MHD simulations.

Neural network approach provides short-term predictions up to 4 hours in the paper [20]. It also experiences significant difficulties predicting sudden commencements of strong geomagnetic storms with $Kp > 5$. Adaptive filtering seems more successful being able to provide 8-hour predictions in the paper [9]. However, in the papers, which incorporate adaptive filtering, the volume of the dataset does not usually exceed 6 months of data (4380 points), which is not enough to correctly describe long-time variations in geomagnetic activity, caused 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 [23]. 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 [3], 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, which based on the Chapman-Ferraro models [26–30] and Dungey model [31]. Global MHD simulations give the longest prediction times but fail to correctly describe kinetics in boundary layers and ballooning instabilities, which are believed to be responsible for substorm onset. Besides, the ring current cannot be described in the framework of ideal MHD.

In this paper I will apply the method, which combines statistical and empirical approaches. I predicted Dst 1 hour ahead because the temporal resolution of the dataset was 1 hour, so I just predicted the next value in the time series. This provided the greatest accuracy and besides, 1 hour is roughly equal to the time, which takes the perturbations of the solar wind to propagate to the Earth. This method appeared to be perfectly suitable for searching for geoeffective parameters. Here, I will try

to extend its predictive capabilities to 3 hours ahead. 3 hours prediction time was selected because on one hand such a forecast is already useful for practical applications and on the other hand, still provides RMS low enough to provide precise forecast.

II. DATA AND ROUTINE

I used the OMNI 2 database, available at NSSDC website [32]. It contains IMF, solar wind and geomagnetic data, averaged over 1-hour intervals (49 parameters in total, starting from Jan 1, 1963). I supplemented it with provisional Dst data, taken from WDC for Geomagnetism (Kyoto) [33]. Thus I obtained a continuous 44-year Dst time series, featuring eye-visible 27-day and 11-year periodicities, which confirms the strong dependence of Dst index on solar activity.

We seek Dst in a regression form

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

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. Naturally, x_i include only quantities measured before the prediction is made. The 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 [34, 35]. Note that notation $x_i(j)$ does not imply that the value x_i is taken at the moment j ; it just indicates that the values of regressors x_i depend on j , though not explicitly.

Regressors were selected using known models of solar wind-magnetosphere interaction with the help of correlation functions; some regressors were selected by trial and error. The initial number of regressors was taken deliberately excessive to let the F-test select the most statistically significant of them.

We estimated geoeffectiveness of the parameter by coefficients and statistical significances of all regressors, which contain this parameter. 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 sample obtained from the initial dataset after rejecting filled values.

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 & Li, 2002), and the linear correlation coefficient (LC) between the prediction and the official Dst index.

cient (LC) between the prediction and the official Dst index.

Note that the described routine should be applied to a large dataset. As an experiment, we applied this routine to datasets for separate years. Each time we obtained different sets of significant regressors, some of which were insignificant neither for other years, nor for the complete dataset. We also encountered situations when a regressor, significant over a complete dataset, suddenly became insignificant for an individual year. We believe these phenomena are connected with individual peculiarities of particular years. Additionally, using a large dataset lifts the problem of determining the number of degrees of freedom for the F-test, because in the case of a large dataset this value can be considered infinite. Another benefit from using a large dataset is good stability of the coefficients C_i — if you build a regression over 22 years of data and use these coefficients to forecast the subsequent 22 years of data, you will obtain the results, nearly as good, as if you were using the whole 44-year dataset for determining the coefficients. So, their characteristic decay time is at least one Hale's cycle. So, after the initial determination of the coefficients one should just calculate a simple polynomial, which takes only a fraction of a second on an average PC, so the large number of regressors should not be a problem.

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

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

where N is the oldest Dst value, and k is the prediction time; 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) for $k = 1$. 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 [23] 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. For $k = 3$, the oldest significant value corresponds to $i = 358$ (14 days and 22 hours).

After determining which previous Dst values are statistically significant, we added all the solar wind parameters available in the OMNI 2 database. Then, we added nonlinear terms. We tried different powers of the most significant terms and their products.

III. RESULTS AND DISCUSSION

For $k = 1$ we obtained 63 significant regressors. The regression has the following characteristics: RMS = 3.76 nT, PE = 0.975, LC = 0.987.

For $k = 3$ we obtained a regression with 178 significant regressors. It has RMS of 7.61 nT, prediction efficiency

of 0.899 and linear correlation with the official Dst of 0.906. The corresponding scatter plot is given on Fig. 1.

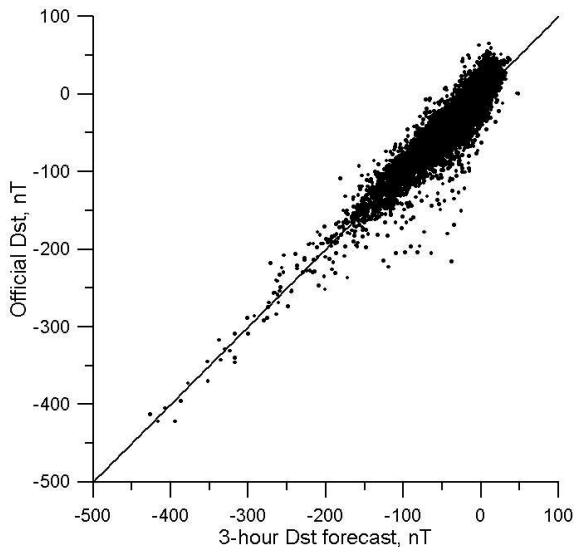


Fig. 1. A scatterplot for the 3-hour prediction over the complete dataset.

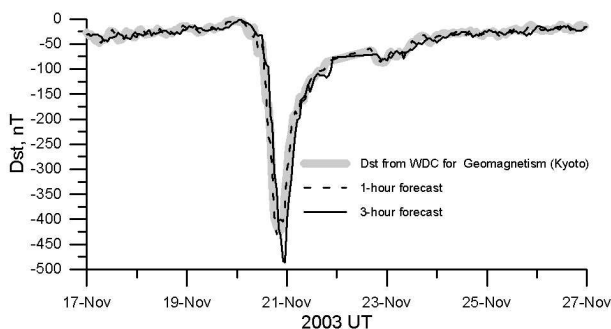


Fig. 2. Actual and predicted Dst plot for the Nov 20, 2003 geomagnetic superstorm.

The most significant regressors are the previous Dst values and various powers of the IMF southern component.

Earlier linear statistical models for Dst prediction suffered from time shift when predicting for more than one hour. Our model is almost free from this artefact due to an extensive use of nonlinear regressors. As an example, we provide Fig. 2, which depicts a major geomagnetic

storm of Nov 20, 2003. A wide solid grey line shows actual Dst measurements, a solid black line shows 3-hour prediction, and a dashed black line shows 1-hour prediction. For comparison, we present Fig. 3, reprinted from the paper by Pallocchia *et al.* [17], who used Artificial Neural Network technique to predict the same event 1 hour ahead.

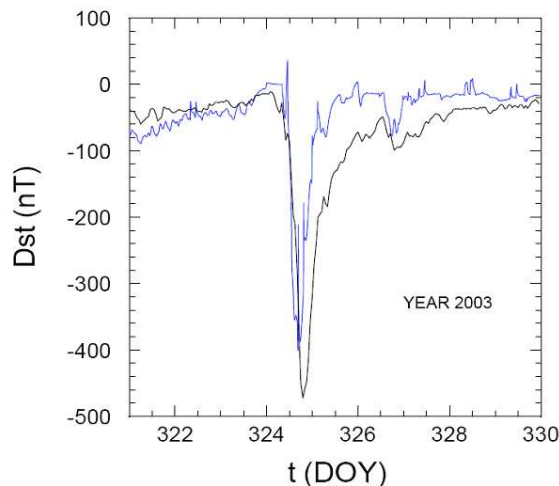


Fig. 3. Predictions of Pallocchia *et al.* [17] for the same event as one on Fig. 2.

IV. SUMMARY

In this paper the prediction time of quite successful statistical Dst forecast technique was extended to 3 hours with reasonable LC and PE values (about 90%). Note that this technique was created with real-time Dst prediction in mind and after the initial determination of the coefficients, which is performed only once, just a simple polynomial function is to be recalculated every hour. This makes this technique especially useful for practical applications.

V. ACKNOWLEDGEMENTS

The author is grateful to the National Space Science Data Center for the OMNI 2 database and to the WDC for Geomagnetism (Kyoto) for Dst index data.

-
- [1] K. Marubashi, *Space Sci. Rev.* **51**, 197 (1989).
 - [2] W. H. Campbell, *J. Atmos. Terr. Phys.* **58**, 1171 (1996).
 - [3] R. K. Burton, R. L. McPherron, C. T. Russell, *J. Geophys. Res.* **80**, 4204 (1975).
 - [4] J. A. Valdivia, A. S. Sharma, K. Papadopoulos, *Geophys. Res. Lett.* **23**, 2899 (1996).
 - [5] S. Kugblenu, S. Taguchi, T. Okuzawa, *Earth Planets Space* **51**, 307 (1999).
 - [6] T. P. O'Brien, R. L. McPherron, *J. Atmos. Sol.-Terr. Phys.* **62**, 1295 (2000).
 - [7] T. P. O'Brien, R. L. McPherron, *J. Geophys. Res.* **105**, 7707 (2000).
 - [8] M. A. Balikhin, O. M. Boaghe, S. A. Billings, H. S. C. K. Alleyne, *Geophys. Res. Lett.* **28**, 1123 (2001).
 - [9] R. F. Harrison, P. M. Drezet, in *Proc. Les Woolliscroft memorial Conf., Sheffield Space Plasma Meeting: Multipoint measurements versus theory. Sheffield, UK, April 24-26, 2001* (ESA SP-492), 141 (2001).

- [10] M. Temerin, Xinlin Li, J. Geophys. Res. **107**, 1472 (2002).
- [11] M. Temerin, Xinlin Li, J. Geophys. Res. **111**, A04221 (2006).
- [12] S. Watanabe, E. Sagawa, K. Ohtaka, H. Shimazu, Earth Planets Space **54**, 1263 (2002).
- [13] P. Ballatore, W. D. Gonzalez, Earth Planets Space **55**, 427 (2003).
- [14] H. L. Wei, S. A. Billings, M. A. Balikhin, Nonlin. Proc. Geophys. **11**, 303 (2004).
- [15] C. Cid, E. Saiz, Y. Cerrato, in *Proc. Solar Wind 11 – SOHO 16 “Connecting Sun and Heliosphere”, Whistler, Canada, 12–17 June, 2005* (ESA SP-592), 116 (2005).
- [16] G. Siscoe, R. L. McPherron, M. W. Liemohn, A. J. Ridley, G. Lu, J. Geophys. Res. **110**, A02215 (2005).
- [17] G. Pallochia, E. Amata, G. Consolini, M. F. Marcucci, I. Bertello, Mem. Soc. Astron. It. Suppl. **9**, 120 (2006).
- [18] X.-Y. Zhou, F.-S. Wei, Earth Planets Space **50**, 839 (1998).
- [19] G. K. Rangarajan, L. M. Barreto, Earth Planets Space **51**, 363 (1999).
- [20] S. Wing *et al.*, J. Geophys. Res. **110**, A04203 (2005).
- [21] X. Li, M. Temerin, D. N. Baker, G. D. Reeves, D. Larson, S. G. Kanekal, Eos **84**, 37 (2003).
- [22] S. Y. Oh, Y. Yi, J. Korean Astron. Soc. **37**, 151 (2004).
- [23] J. R. Johnson, S. Wing, Report PPPL-3919rev, http://www.pppl.gov/pub_report/2004/PPPL-3919rev.pdf.
- [24] M. Dryer, S. T. Wu, G. Gislason, S. M. Han, Z. K. Smith, J. F. Wang, D. F. Smart, M. A. Shea, Astrophys. Space Sci. **105**, 187 (1984).
- [25] J. Raeder *et al.*, J. Geophys. Res. **106**, 381 (2001).
- [26] S. Chapman, V. C. A. Ferraro, Terr. Magn. Atmos. Electr. **36**, 77 (1931).
- [27] S. Chapman, V. C. A. Ferraro, Terr. Magn. Atmos. Electr. **36**, 171 (1931).
- [28] S. Chapman, V. C. A. Ferraro, Terr. Magn. Atmos. Electr. **37**, 147 (1932).
- [29] S. Chapman, V. C. A. Ferraro, Terr. Magn. Atmos. Electr. **38**, 79 (1933).
- [30] S. Chapman, V. C. A. Ferraro, Terr. Magn. Atmos. Electr. **45**, 245 (1940).
- [31] J. W. Dungey, Planet. Space Sci. **10**, 233 (1963).
- [32] OMNI 2 database, <http://nssdc.gsfc.nasa.gov/omniweb/>
- [33] Kyoto World Data Center for Geomagnetism, <http://swdcdb.kugi.kyoto-u.ac.jp/>
- [34] R. A. Fisher, *Statistical methods for research workers* (London, Oliver and Boyd, 1954).
- [35] D. J. Hudson, *Statistics Lectures on Elementary Statistics and Probability* (Geneva, CERN, 1964).

СТАТИСТИЧНИЙ ПІДХІД ДО ПЕРЕДБАЧЕННЯ Dst

А. С. Парновський

*Інститут космічних досліджень НАНУ і НКАУ,
просп. акад. Глушкова, 40, Київ-187, 03680, Україна*

Використовуючи статистичний підхід, створено феноменологічну модель взаємодії сонячного вітру з магнітосферою. Ця модель дає змогу вимірювати відносну геоефективність різних фізичних параметрів, виявляти деякі емпіричні залежності, робити точний прогноз геомагнітної активності до 3 годин уперед та оцінювати геомагнітну активність до 9 годин уперед. При прогнозуванні на 1 годину забезпечено коефіцієнт кореляції $LC = 0.987$ та ефективність прогнозу $PE = 0.975$. При прогнозуванні на 3 години — $LC = 0.906$ та $PE = 0.899$. При прогнозуванні на 9 годин — $LC = 0.820$ та $PE = 0.711$.