How good is the prediction of space weather based on solar and interplanetary properties?

N. Srivastava

Udaipur Solar Observatory, Physical Research Laboratory, P.O. Box 198, Badi Road, Dewali, Udaipur, India

Abstract. Space weather prediction involves advance forecasting of the magnitude and the on-set time of major geomagnetic storms at the earth. In a previous attempt, a logistic regression model based on solar and interplanetary variables was developed which was based on an exhaustive study of the solar origins of major geomagnetic storms recorded during 1996-2002. In this paper, the logistic regression model developed earlier is refined using a new database of the solar and interplanetary characteristics of the major geomagnetic storms recorded during 2003-2004, which leads to better prediction results. The model is also used to estimate the relative importance of each solar and interplanetary variable in predicting major geomagnetic storms. In an attempt aimed at an early prediction of the occurrence of geomagnetic storms, the interplanetary variables are excluded and a new model based only on the solar variables is developed. The new model did not perform well, which indicates that the solar variables responsible for geomagnetic activity at the Earth are not well-understood.

Index Terms. Coronal mass ejections, geomagnetic storms, interplanetary properties, space weather prediction.

1. Introduction

Space weather prediction involves the integration of four major areas of the solar terrestrial relations, namely, solar physics, interplanetary physics, magnetospheric physics and ionospheric physics. For a successful quantitative prediction of space weather it is required to achieve capability of accurate prediction of the magnitude of the resulting storm and its arrival time at the earth. This capability comes from successful

(a) identification of solar sources or origins namely the coronal mass ejections or CMEs.
(b) understanding of propagation of CMEs in the interplanetary medium.
(c) Estimation of key interplanetary parameters e.g., solar wind velocity, solar wind density, the southward component of the interplanetary magnetic field and the total interplanetary magnetic field.
(d) understanding of the physical relationship of interplanetary parameters with solar parameters.
(e) understanding of the relationship of the strength of the geomagnetic storm on the interplanetary parameters.
(f) study of the effects of geomagnetic storms on navigation system, satellite communications, etc.

Several attempts have been made till now to address the above jointly or separately, since the first attempt made by Chapman and Bartels (1940) and Snyder et al. (1963). Even with the new understanding, the prediction scheme is not very reliable, although with the capability of measuring the speeds of the CMEs, one is now able to predict the transit time with a much better accuracy than before. A few recent studies aimed towards understanding and improving space weather prediction schemes are Cane et al. (2000), Feynman and Gabriel (2000), Gopalswamy (2000; 2001a; 2001b), Wang et al. (2002), Zhang et al. (2003), Srivastava and Venkatakrishnan (2004). These studies have also led to development of empirical models to predict the arrival time of the CME at the earth, accurately.

2. Observational inputs for space weather prediction

(a) Key solar parameters: Recent studies owing to highly sensitive observations recorded by EIT (Delaboudienere et al., 1995) and LASCO (Brueckner et al., 1995) aboard SOHO, have clearly demonstrated that CME morphological properties, viz., their association with halos or partial halos can be used as a characteristic of a geo-effective event. Further, the location of the CME on the solar surface expressed in terms of heliographic latitude and longitude, its association with flare/eruptive prominence and source active region properties viz. magnetic energy are also important solar parameters representative of geo-effectiveness. (Srivastava and Venkatakrishnan, 2002; 2004; Wang et al., 2002; Zhang et al., 2003; Schwenn et al., 2005).

(b) Key interplanetary parameters: The main inputs for any space weather prediction based on properties of interplanetary medium are known to be the solar wind speed at 1 AU, the total interplanetary magnetic field (B_T) and the southward component of the interplanetary magnetic field (B_z). Some other well known interplanetary parameters are the interplanetary shock speed which is dependent on the speeds and the densities before and after the shock, the ram pressure, which is the pressure exerted by the solar wind on the magnetosphere (Srivastava et al., 2004).

However, it was first shown by Burton et al. (1975) that there exists an empirical dependence of D_{ST} index on two main parameters of the solar wind, the solar wind speed and the southward component of the IMF. Several real-time prediction schemes have been worked out based on original Burton's formula for example, that by O'Brien and McPherron (2000), Feldstein (1992), Wu and Lundstedt (1996), and Fenrich and Luhmann (1998). These schemes being real-time prediction schemes are largely based on inputs from the *in-situ* measurements of the solar wind, therefore the forewarning of the geomagnetic storms can be made only 45 minutes to an hour prior to the actual commencement of the storm. For an early prediction of a geomagnetic storm, it is required that the crucial interplanetary parameters be predicted well ahead in time. For this purpose, one needs to identify the important solar parameters and their influence on the interplanetary parameters.

After the launch of SOHO, several prediction schemes based on halo CMEs as inputs have been attempted since LASCO aboard SOHO regularly recorded halo CMEs. However, these schemes failed to predict a significant proportion of magnetic storms (about 20%) which fall in the category of "missing alarms". Similarly, 15% of the predicted events never occur and fall in the class of "false alarms" (Schwenn et al., 2005).

3. Recent approaches in space weather prediction

The modern schemes for the prediction of space weather include multi-directional approaches all leading to the ultimate goal of predictive capability of the arrival time of the CME at the earth and the magnitude of the resulting geomagnetic storms. These approaches or schemes may involve development of either (a) statistical/empirical/semi-empirical (b) physics based or (c) neural network models. Some of the prediction schemes currently being used are described below.

3.1 Arrival-time prediction model

Various empirical models have been developed to estimate the arrival time of the CMEs at the earth. These are based on the CME observations made close to the sun. Normally halo or partial halo CMEs are known to be associated with geo-effectiveness at the earth. Using different measurable properties of such CMEs, the following empirical models are currently used to predict the arrival time. These include

(a) Schwenn model (2000) which uses both plane of sky speed as well as expansion speed measured in the perpendicular direction and requires observations of a CME in LASCO-C3**.**

(b) Gopalswamy et al. (2001) model which uses fastest plane-of-sky speeds for prediction of arrival time of ICMEs or magnetic clouds.

(c) Smith model (2003) which uses halo CMEs and its properties as inputs. It also depends on the location of the source and the fastest CME speed measured in the plane-of-sky.

3.2 Semi-empirical CME models for space weather prediction

(a) Solar magnetogram based model developed by Wang and Sheeley (1990) which does not include any transients, instead solar wind structures are predicted on the basis of photospheric magnetograms. This model also includes high-speed/low-speed solar wind stream interactions.

(b) Opening coronal field model developed by Luhmann et al. (1998) is based on the assumption that the CMEs are associated with solar magnetic field changes taking place on the photosphere. This model uses the coronal PFSS (Potential Field Source Surface Model) to observe the coronal magnetic field changes.

3.3 Physics based models

Dryer (1994) modeled fast propagating CMEs and the associated shocks through the heliosphere, using 1 or 2 dimensional MHD codes for prediction of the interplanetary shock speeds and the arrival time at the earth. Odstrcil and Pizzo (1999) modified it to 3-dimensional model and also included ambient solar wind parameters and the structure of the heliospheric current sheets. Both the models are useful for near real time prediction based on a detection of a metric type II shock which is used as input along with other parameters like, GOES X-ray flare association and its location on the solar surface.

3.4 Neural networks models

This model developed by Lundstedt et al. (2002) consists of a recurrent neural network that requires hourly averages of the solar wind magnetic field component B_z, particle density n, and velocity V as inputs and predicts the D_{ST} index in almost real-time.

4. Our approach: Logistic regression model development

Our effort for space weather prediction is based on an observational approach involving the following steps:

(1) study statistical properties of solar and interplanetary sources of geo-effective events recorded during 1996-2002;
(2) identify key solar and interplanetary parameters which influence geo-effectiveness from (1);
(3) develop a simple logistic regression model using the key parameters, solar and interplanetary, identified with the occurrence or non-occurrence of an intense storm;
(4) Validation of the model to test its predictive capability.

Following this approach, we developed a logistic regression model based on observations of 46 geo-effective CMEs observed during 1996-2002 (Srivastava, 2005a). In this model, the storms with $-200< D_{ST}<-100$ nT were classified as "intense" and other events with $D_{ST} <-200$ nT as "super-intense". The classification is such that at least a sufficient number of geo-effective events fall in both the categories and thus a successful prediction model can be estimated.

In this space weather prediction model, a binary dependent variable representing the occurrence of intense/super-intense storms is regressed against a series of independent model variables defined by a number of solar and interplanetary properties and have been described in detail in Srivastava and Venkatakrishnan (2004).

Mathematically, the logistic regression equation is given by

$$P=1/(1+e^{-Z})$$

$$\text{Where, } Z=b_0+b_1\,x_{i1}+\ldots\ldots\ldots b_j x_{ij}$$

Here P is the probability of occurrence of intense or super-intense geomagnetic storm given the observation of solar and interplanetary parameters. The Z_i is the value of continuous variable. The details of the model and variables are described in Srivastava (2005a).

4.1 Data sets used and model estimation

In the present study, we extended the previous model to a larger database which included the geo-effective events of 2003-2004. This increased the total number of events for the training data sets from 46 to 55. This extended data base included the super-intense geomagnetic storms of the current solar cycle which occurred during October-November, 2003. The estimated logistic regression model was tested for its predictive capability of occurrence of major geomagnetic storms using only solar parameters as input.

Table 1 shows, that the data set used in this paper included a total of 67 events, out of which 55 events were used as training and 12 events as validation data set. These data sets were selected such that the training data set included about 20 super-intense and 35 intense events. On the other hand, the validation data set included 7 super-intense and 5 intense events. A logistic regression model was estimated using the same solar and interplanetary parameters as inputs as in previous study (Srivastava, 2005a). This equation is given by

$$D_{ST-B}=1/(1+\exp(-(-5.23-0.12xV_{sh-B}+3913474.8xP_R+0.83x\,H_B-4.9x10^{-2}xFL_B+0.36xL_B+8.3x10^{-4}xV_i-3.3x10^{-2}xB_T+0.19xB_z))) \quad (1)$$

Here, D_{ST-B} denotes the probability of occurrence of intense storms, expressed as a function of several solar and interplanetary inputs. In particular, V_{SH-B} denotes the coded values for shock speed, P_R denotes the ram pressure exerted on the magnetosphere by the solar wind, V_i is the initial speed of the CME in the plane-of-sky, H_B denotes the binary variable for association with partial or full halo, and FL_B specifies the binary variable for association of the CME with flare or eruptive prominence. L_B denotes the binary variable for the source location of the CME within and B_z and B_T the southward component and the total interplanetary magnetic field respectively.

Table 1. Prediction with Solar and Interplanetary variables

Data sets		Observed	Predicted	% Correct Prediction
Training	Super-intense	20	13	65
	Intense	35	33	94
	Total	55	46	84
Validation	Super-intense	7	5	71
	Intense	5	4	80
	Total	12	9	75

5. Results

As mentioned earlier, most of the current prediction models depend on the *in-situ* measured values of interplanetary parameters and therefore give only one hour advance warning of the occurrence of a storm. However, for achieving reliable prediction soon after a CME is launched in the direction of the earth, it is not only essential to identify the key solar parameters which influence the well known interplanetary parameters that produce major storms, but also investigate the relative importance of various solar parameters that have been used as inputs to the model. To achieve this goal of early prediction of the geomagnetic storms, we considered only solar parameters as inputs in this data set. The logistic regression model was estimated using the training data set of 55 geo-effective events recorded during 1996-2002 and is given by the following equation:

$$D_{ST-B}=1/(1+\exp(-(-2.97+0.83xH_B-0.21xFL_B+1.05xL_B+1.1x10^{-3}xV_i))) \quad (2)$$

Here, the selected solar parameters include binary variables for the association of sources with halos, flares and the location of the source regions respectively, and the initial speed of the coronal mass ejections as discussed in the previous section. The equation of the model is then used for validation tests on the 12 geoeffective events recorded during 2003-2004. Any logistic regression model also gives standardized coefficients which may be used to compare the relative weights of the variables. The higher the absolute value of a coefficient, the more important is the weight of the corresponding variable, in the model. When the confidence interval around standardized coefficients has a value 0, the weight of a variable in the model is not considered significant.

Fig.1 shows standardized coefficients representing the relative weights of various input variables to the model. It is clearly seen that amongst the chosen input solar parameters, the association with full halos is the most important, followed by the initial speeds of these halos, location of the halo CMEs and their association with flares.

Table 2. Prediction with Solar Variables Only

Data sets		Observed	Predicted	% Correct Prediction
Training	Super-intense	20	11	55
	Intense	35	31	88
	Total	55	42	76
Validation	Super-intense	7	4	57
	Intense	5	4	80
	Total	12	8	75

The Table 1 shows the prediction capability of the model for both the training (55 events) and validation (data sets). The 4th Column of Table 1 and Table 2 show correct prediction percentage for the model using (a) both solar and interplanetary variables (b) solar variables only, respectively. It can be clearly noticed that while 84% of the training data set could be correctly predicted using both solar and interplanetary parameters, the percentage of correct prediction for the training data set reduces to 76% if only solar parameters are used as input to the model.

Similarly 75% of the validation data set could be correctly predicted by the estimated model equation while approximately 66% of the events, for which only solar parameters were used, were correctly classified by the model.

We compared our results for the geo-effective events recorded during 1996-2004 with the previous work which was limited up to 2002. The classification was done correctly for 62.5% of the super-intense storms and 97% of the intense geomagnetic storms from the training data set. The current study shows that model correctly classifies 65% of the super-intense storms and 94% of the super-intense storms from the training data set. This result is more or less similar to the previous result.

However, the improved forecasting capability of the model becomes obvious if one compares the results of validation tests for the data set up to 2002 with the current data set extended up to 2004. While only 50% of the super-intense geomagnetic storm of the validation data set were correctly classified in the previous study, the results of the study show that 71% of the super-intense storms could be classified correctly. This is due to the fact that more number of super-intense events, which are relatively less frequent, were available in the current data set than used in the previous study. These included superstorms of October 28 and 29, 2003 and also of November 20, 2003. The latter being the strongest storm of the current solar cycle (Srivastava, 2005b). These super-storms (D_{ST} <-300 nT) were correctly predicted by the present model.

Our results also indicate that the forecasting capability specially for the super- intense events is poorer, i.e., decreases from 71% to 57%, if only solar parameters are used as input to the logistic regression model. However, the percentage of correct prediction for intense events is approximately the same about ~80%.

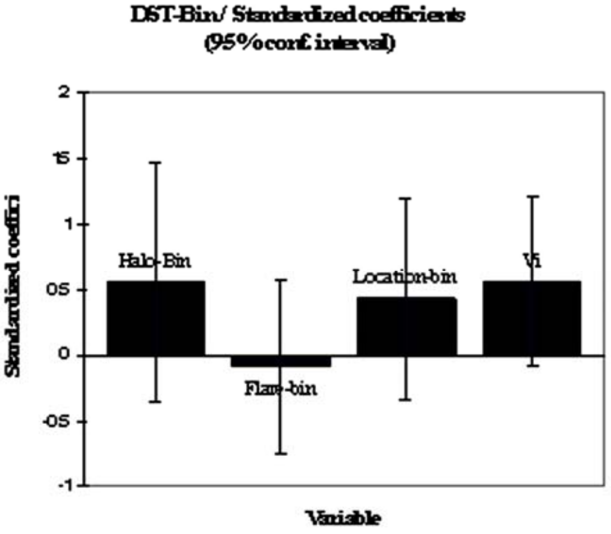

Fig. 1. Standardized coefficients of the solar parameters used as input to the predictive model.

6. Conclusion

In the present study we estimated a simple logistic regression model for predicting the occurrence of intense/super-intense geomagnetic storms based on a number of solar and interplanetary variables. This model was estimated from a database of 55 geo-effective events recorded during 1996-2002 by using (a) both solar and interplanetary parameters (b) solar parameters only and validated on 12 geo-effective events recorded during 2003-2004.

From this study we conclude the following:

(1) Results from regression model based on interplanetary and solar variables

Our study shows that the capability of prediction of the occurrence of geomagnetic storms improves for super-intense storms with a larger database. As the number of super-intense events which occurred during 2003-2004 increased, it resulted in the increased efficiency of prediction.

The model is highly successful (80%) in predicting the occurrence of intense geomagnetic storms and moderately successful (70%) in predicting super-intense storms. The model also indicates two of the interplanetary parameters for example, the southward component of the interplanetary magnetic field, B_z and total interplanetary magnetic field B_T are the most important parameters for prediction. This result is in accordance with the results of previous authors.

Amongst solar parameters, initial speed of the coronal mass ejections, V_i is the most important parameter followed by association with full halo, and location and association with flares or eruptive prominences.

(2) Results from regression model based on solar variables only

The logistic regression model based on solar parameters only shows poor predictive capability for the super-intense events. The test shows that the chosen parameters for the purpose of space weather forecasting may not be sufficient. This implies that besides V_i, other solar parameters like magnetic field magnitude and its orientation on the sun, should also be considered for example as shown by Gopalswamy et al. (2005b) for the CME of November 18, 2003, which gave rise to the strongest storm of the current cycle.

(3) Future work

The present model cannot be used to predict the exact magnitude of the resulting geomagnetic storm therefore the future work will be aimed towards this direction. Further, to improve the advance forecasting based solely on solar parameters, additional solar parameters i.e. magnetic field signatures on the sun will be included in the present model, while the existing parameters for example, the initial speeds will be refined further for improving the forecasting capability.

Acknowledgments. The author acknowledges the LASCO and CELIAS operations teams on SOHO and ACE operation teams for the data which was used in the current study. The author would also like to thank the organizers of the ILWS for providing the local hospitality to attend the workshop.

References

G. E. Brueckner et al., “The large angle spectroscopic coronagraph (LASCO)”, Solar *Phys*., vol. 162, pp 357-402, 1995.

R. K. Burton, R. L. McPherron and C. T. Russell, “An empirical relationship between interplanetary conditions and D_{ST}”, *J. Geophys. Res.*, vol. 80, pp. 4204-4214, 1975.

H. V. Cane, I. G. Richardson and O.C. St. Cyr., “Coronal mass ejections interplanetary ejecta and geomagnetic storms”, *Geophys. Res. Lett*., vol. 27, pp. 3591-3594, 2000.

S. Chapman and J. Bartels, *Geomagnetism*, Oxford, Oxford University Press, 1940.

J. -P. Delaboudiniere et al., “Extreme ultraviolet imaging telescope for the SOHO mission”, *Solar Phys*., vol. 162, pp. 291-312, 1995.

M. Dryer, “Interplanetary Studies: Propagation of disturbances between the sun and the magnetosphere”, *Space Sci. Rev.*, vol. 67, pp. 363-419, 1994.

Y. I. Feldstein, “Modelling of the magnetic field”, *Space Sci. Rev*., vol. 59, pp. 83-165, 1992.

F. R. Fenrich and J. G. Luhmann, “Geomagnetic response to magnetic clouds of different polarity”, *Geophys. Res. Lett*,. vol. 25, pp. 2999-3002, 1998.

J. Feynman and S. B. Gabriel, “On space weather consequences and predictions”, *J. Geophys. Res.*, vol. 105, pp. 10543-10564, 2000.

N. Gopalswamy, A. Lara, R. P. Lepping, M. L. Kaiser, D. Berdichevsky and O. C. St. Cyr, “Interplanetary acceleration of coronal mass ejections”, *Geophys. Res. Lett.*, vol. 27, pp. 145-148, 2000.

N. Gopalswamy, A. Lara, S. Yashiro, M. L. Kaiser and R. A. Howard, “Predicting the 1-AU arrival times of coronal mass ejections”, *J. Geophys. Res.*, vol. 106, pp. 29207-29218, 2001a.

N. Gopalswamy, A. Lara, M. L. Kaiser and J. -L. Bougeret, “Near – sun and near Earth manifestation of solar eruptions”, *J. Geophys. Res*., vol. 106, pp. 25261-25278, 2001b.

N. Gopalswamy, S. Yashiro, Y. Liu., G. Michalek, A. Vourlidas, M. L. Kaiser and R. A., Howard, “Coronal mass ejection and other extreme characteristics of the 2003 Oct-Nov solar eruptions”, *J. Geophys. Res*, vol. 110, doi:10.1029/ 2004JA010958, 2005a.

N. Gopalswamy, S. Yashiro, G. Michalek, H. Xie, R. P. Lepping and R. A. Howard, “Solar source of largest geomagnetic storm of cycle 23”, *Geophys. Res. Lett.*, vol. 32, doi:10.1029/ 2004GL021639, 2005b.

J. G. Luhmann, J. T. Gosling, T. Hoeksema., X. Zhao, “The relationship between large-scale solar magnetic field evolution and coronal mass ejections”, *J. Geophys. Res*, vol. 103, pp. 6585-6594, 1998.

H. Lundstedt, H. Gleisner and P. Wintoft, “Operational forecasts of the geomagnetic Dst index” , *Geophys. Res. Lett.*, vol. 29, pp. 34(1-4), doi:10.1029/2002GL016151.

T. P. O'Brien and R. L. McPherron., “An empirical phase space analysis of ring current dynamics: Solar wind control of injection and decay”, *J. Geophys. Res*., vol. 105, pp. 7707-7720, 2000.

D. Odstrcil and V. J. Pizzo, “Three-dimensional propagation of CMEs in a structured solar wind flow: 1. CME launched within the streamer belt”, *J. Geophys. Res.*, vol. 104, pp. 493-504, 1999.

R. Schwenn, “Heliospheric 3d Structure and CME Propagation as Seen from SOHO: Recent Lessons for Space Weather Predictions”, *Adv. Space Res.*, vol. 26, pp. 43-53, 2000.

R. Schwenn, A. Dal Lago, E. Huttunen and W. D. Gonzalez, “The association of CMEs with their effects near the Earth”, *Ann. Geophysicae.*, vol. 23, pp.1033-1059, 2005.

Z. Smith, W. Murtagh, T. Detman, M. Dryer, C. D. Fry and C. -C. Wu, “Study of solar-based inputs into space weather models that predict interplanetary shock-arrivals at earth”, in *Solar variability as an input to the Earth’s environment International Solar Cycle studies (ISCS) Symposium*, A. Wilson. Ed. ESA SP-535, Noordwijk: ESA Publications Divisions, 2003, pp. 547-552.

C. W. Snyder, M. Neugebauer, M and U. R. Rao, “The solar wind velocity and its correlation with cosmic-ray variations and with solar and geomagnetic activity”, *J. Geophys. Res*., vol, 68, pp. 6361-6370, 1963.

N. Srivastava, “Predicting the occurrence of super-storms”, *Ann. Geophysicae*, vol. 23, pp. 2989-2995, 2005a.

N. Srivastava, “A logistic regression model for predicting the occurrence of intense geomagnetic storms”, *Ann. Geophysicae*, vol. 23, pp. 2969-2974, 2005b

N. Srivastava and P. Venkatakrishnan, “Relationship between CME Speed and Geomagnetic Storm Intensity”, *Geophys. Res. Lett*., vol. 29, pp. 1287-1290, 2002.

N. Srivastava and P. Venkatakrishnan, “Solar and interplanetary sources of major geomagnetic storms during 1996-2002”, *J. Geophys. Res.*, vol. 109, A010103, doi10.1029/2003JA010175, 2004.

Y. -M Wang, N. R. Sheeley Jr., “Solar wind speed and coronal flux-tube expansion”, *Astrophys. J.*, vol. 355, pp. 726-732, 1990.

Y. -M. Wang, P. Z. Ye, S. Wang, G. P. Zhou and J. X. Wang, “A statistical study on the geo-effectiveness of earth-directed coronal mass ejections from March 1997 to December 2000”, *J. Geophys. Res.*, vol.107, pp.1340, doi:10.1029/2002JA009244, 2002.

J. G. Wu and H. Lundstedt, “Prediction of geomagnetic storms from solar wind data using Elman recurrent neural networks”, *Geophys. Res. Lett*., vol. 23, pp. 319-322, 1996.

J. Zhang, K. P. Dere, R. A. Howard and V. Bothmer, “Identification of solar sources of major geomagnetic storms between 1996 and 2000”, *Astrophys. J.*, vol. 82, pp. 520-533, 2003.