J. Astrophys. Astr. (2006) **27,** 237–242

The Challenge of Predicting the Occurrence of Intense Storms

Nandita Srivastava
Udaipur Solar Observatory, P. O. Box 198, Badi Road, Udaipur 313 001, India.
e-mail: nandita@prl.res.in

Abstract. Geomagnetic super-storms of October and November 2003 are compared in order to identify solar and interplanetary variables that influence the magnitude of geomagnetic storms. Although these super-storms ($D_{ST} < -300\,nT$) are associated with high speed CMEs, their D_{ST} indices show large variation. The most intense storm of November 20, 2003 ($D_{ST} \sim -472\,nT$) had its source in a comparatively small active region and was associated with a relatively weaker, M-class flare, while the others had their origins in large active regions and were associated with strong X-class flares. An attempt has been made to implement a logistic regression model for the prediction of the occurrence of intense/super-intense geomagnetic storms. The model parameters (regression coefficients) were estimated from a training data-set extracted from a data-set of 64 geo-effective CMEs observed during 1996–2002. The results indicate that logistic regression models can be effectively used for predicting the occurrence of major geomagnetic storms from a set of solar and interplanetary factors. The model validation shows that 100% of the intense storms ($-200\,nT < D_{ST} < -100\,nT$) and only 50% of the super-intense ($D_{ST} < -200\,nT$) storms could be correctly predicted.

Key words. Sun: coronal mass ejection—geomagnetic storms—space weather.

1. Introduction

Space weather prediction involves forecasting of the magnitude and the time of the commencement of a geomagnetic storm, based on solar and interplanetary observations. Most of the currently used prediction schemes are based on the formula of Burton *et al.* (1975). They are generally reliable, even though they depend solely on interplanetary (IP) parameters, viz., the solar wind speed and the southward component of the interplanetary magnetic field. However, when solar inputs are used for prediction, one encounters the problem of 'false alarms'(the predicted events never occur) and of 'missing alarms' (no obvious solar signatures of the geomagnetic storm) as also reported by Schwenn *et al.* (2005). Although the presently available prediction schemes accurately predict the D_{ST} index of the geomagnetic storm, their prior warnings are only a few hours ahead of the commencement of the geomagnetic storm. This is because they are based on *in-situ* properties of the solar wind that can only be measured close to the Earth. For example, the ACE spacecraft measurements give about 30 to 60 minutes of warning time as it measures the solar wind properties at the L1 point. A longer warning time requires a solar wind monitor further upstream (McPherron *et al.* 2004). For advance prediction of the arrival time of a CME at the Earth, it is necessary to

- identify the solar origins of geomagnetic storms and
- understand their relation with interplanetary parameters, viz., solar wind speed and the B_z component of the IMF that are responsible for producing super-storms.

To achieve this goal, it is also important to develop a prediction scheme based on both solar and IP properties of geo-effective CMEs.

In the present paper, key solar and IP parameters which may have been responsible for producing super-storms ($D_{ST} < -300$ nT) during October–November 2003 are described. Using these parameters for a larger data-set of 64 geo-effective CMEs recorded during 1996–2002, a logistic regression model was developed to forecast the occurrence of an intense or super-intense storm. For the development of the prediction model, the geomagnetic storms with $D_{ST} < -200$ nT were considered as 'super-intense' while the storms with -200 nT $< D_{ST} < -100$ nT were considered as 'intense'.

2. Super-intense storms of the current solar cycle

During October–November 2003, several powerful CMEs occurred on the Sun which led to intense/super-intense geomagnetic storms (Gopalswamy *et al.* 2005a, b; Srivastava 2005a). Out of these, three super-intense storms that were associated with CMEs and active region flares are described here in Table 1. Although these CMEs emanated at very high speeds for instance, ~ 1175 km s^{-1} (November 18, 2003) and 2125 km s^{-1} (October 28, 2003), the source active regions of these CMEs showed a remarkable difference in area. The November 18 CME originated from a smaller region and was associated with a smaller and weaker flare (M3.9, 2N) as compared to the CMEs of October 28 and 29, both of which originated from a larger active region (Figs. 1 and 2). All the three geo-effective CMEs gave rise to strong interplanetary

Table 1. Source regions of super-storms during October–November 2003.

Date	October 28, 2003	October 29, 2003	November 18, 2003
Source flare	X17.2 4B flare in NOAA AR 10486 at 10:50 UT	X10.2 2B flare in NOAA AR 10486 at 20:50 UT	M3.9, 2N flare in NOAA AR 10501 at 8:30 UT
Active region area (corrected hemispherical area)	2500×10^{-6} $A_\odot$	2500×10^{-6} $A_\odot$	370×10^{-6} $A_\odot$
Speed in LASCO-C2	2125 km s^{-1}	1950 km s^{-1}	1175 km s^{-1}
IP shock speed	$\sim$2000 km s^{-1}	1000 km s^{-1}	$\sim$730 km s^{-1}
D_{ST} index	-363 nT	-400 nT	-472 nT

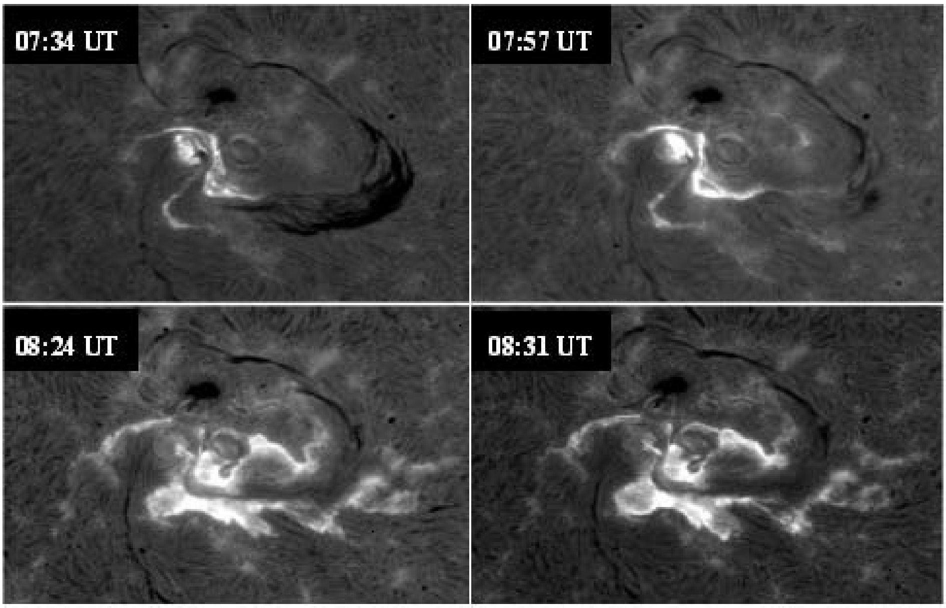

Figure 1. Time-lapse images of an M3.9, 2N flare recorded in H-alpha at Udaipur Solar Observatory. This flare was associated with the large CME of November 18, 2003.

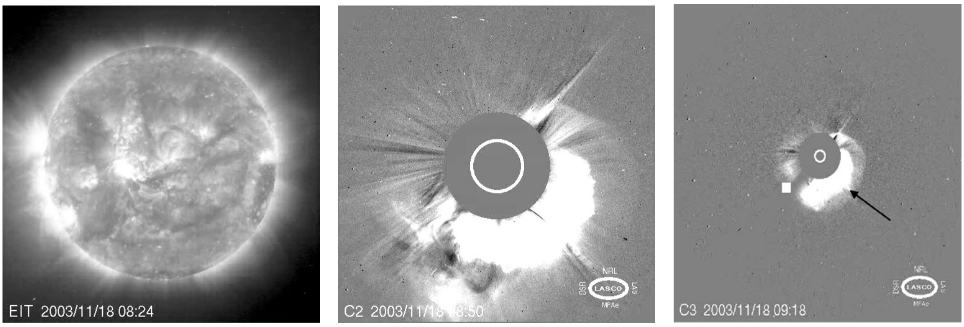

Figure 2. CME of November 18, 2003 which was the source of the strongest storm of the current solar cycle. The first image shows the flare as observed in EIT (195 Å) on SOHO. The middle and right images are recorded in white light by LASCO-C2 and C3.

shocks which were recorded by CELIAS aboard SOHO (http://umtof.umd.edu/pm). The corresponding IP shock velocities ranged between 730 and 2000 km s^{-1}. A deceleration of ~ -3.3 m s^{-2} was observed for the CME of November 18, 2003 as it travelled through the IP medium. On the other hand, the CME of October 28, 2003 is an example of faster CME with a relatively lower value of deceleration ~ -1.8 m s^{-2}. The arrival time of November 18 CME was $\sim$48 hours as compared to 19 hours in the case of October 28 CME, nevertheless, the magnitude of the resulting storm was greater in the case of the former making it the strongest storm of the current solar cycle ($D_{ST} \sim -472$ nT). Although our previous studies indicate that the initial speed of a CME is one of the key factors that dictate the magnitude of the resulting geomagnetic storm (Srivastava & Venkatakrishnan 2002, 2004), the November 18 event is an exception in that, it had a relatively low initial speed. However, the CMEs of October 28 and 29 had high initial speeds. This further suggests that there may be other solar factors that influence the magnitude of the resulting geomagnetic storms.

With exception of the November 18, 2003 CME, the initial speed of a CME may be considered as the most reliable predictor (amongst solar variables) of the strength of the associated geomagnetic storm because fast mass ejections are responsible for building up the ram pressure at the earth's magnetosphere. This result is based on a comparative study of these events with other super-storms of the current solar cycle (Srivastava 2005a). This study also showed the following:

- If the source active region of a CME is large in area, it posseses a large amount of magnetic energy. The larger the magnetic energy available, the higher the speed of the ensuing CME which causes higher ram pressure of the solar wind on the magnetosphere and compression of the southward component of the IMF (B_z).
- The strength of a geomagnetic storm not only depends on the interplanetary-magnetospheric coupling parameter VB_z, but also on the duration (T_{B_z}) of B_z. The higher the value of $VB_zT_{B_z}$ the higher will be the strength of the storm.

3. Space weather prediction: Development of logistic regression model

Solar and interplanetary parameters as described in the previous section were specified for 64 geo-effective CMEs observed during 1996–2002 and were used to develop a logistic regression model to predict the occurrence/non-occurrence of a severe storm. The geomagnetic storms with $-200 < D_{\rm ST} < -100$ nT were classified as 'intense' and other events with $D_{\rm ST} < -200$ nT as 'super-intense'. It may also be pointed out here, that the 'super-storms' with $D_{\rm ST} < -300$ nT as observed during October–November 2003, are very rare. Therefore, the above classification was necessary, in order to have at least a sufficient number of events in both the categories so that a successful predictive model can be developed.

In the logistic regression model, a binary dependent variable indicating the occurrence of an intense or super-intense geomagnetic storms is regressed against a series of independent model variables that define a number of solar and interplanetary properties of the 64 geo-effective CMEs. The input variables for these models were chosen from Srivastava & Venkatakrishnan (2004). The value of the independent variable was taken as 1 for super-intense storms ($D_{\rm ST} < -200$ nT) and 0 for intense storms ($-200 < D_{\rm ST} < -100$ nT). The logistic regression equation is:

$$P = \frac{1}{(1 + e^{-Z})}, \tag{1}$$

where

$$Z = b_0 + b_1 x_{i1} + \cdots + b_j x_{ij},$$

P is the probability of the occurrence of an intense/super-intense geomagnetic storm, given the ith observation of the solar and interplanetary variables, and Z_i is the value of the continuous variable Z. The main solar inputs to the model included:

- latitude and longitude of the origin of the CME,
- flare/prominence association,

- association with full/partial/non-halo CME and
- initial speeds of the CME (V_i).

The interplanetary inputs were:

- shock speeds (V_{SH}),
- ram pressure (P_R),
- total value of the IMF (B_T),
- southward component of the IMF (B_z),
- solar wind speeds (V_1 and V_2) and densities (n_1 and n_2) before and after the shock and
- solar wind-magnetospheric coupling parameter (VB_z).

The details of the model have been described in Srivastava (2005b). The training of the model comprised estimation of regression coefficients using an iterative maximum likelihood method on a training sub-set of 55 geo-effective events. The trained logistic regression model was validated by predicting the occurrence of intense/super-intense geomagnetic storms for the validation sub-set of 9 events which were not used for training the model. The results show that the logistic regression model correctly classifies 62.5% of the training super-intense storms and 97% of the intense storms. Amongst the validation events, the model correctly classifies 50% of the super-intense storms and 100% of the intense storms.

4. Summary and conclusion

The major geomagnetic storms are mostly associated with CMEs accompanied by flares originating from large active regions located close to the central meridian and at low latitudes. Although the speed of the November 18, 2003 CME was relatively small, the resulting storm of November 20 possibly owes its large magnitude to the long duration for which the B_z component remained southward and to the large magnitude of B_z resulting from the high inclination of the magnetic cloud to the ecliptic plane. Further, if the source active region of a CME is large in area, it possesses a large amount of magnetic energy and hence is capable of producing a high speed CME. The high speed CME is responsible for causing higher ram pressure of the solar wind on the magnetosphere and compression of the southward component of the IMF (B_z). However, it does not imply that if the area is smaller then the ensuing CME cannot be strongly geo-effective, for example, the November 18, 2003 case shows that such CMEs may also lead to severe storms and arrive without prior warning. The strength of a geomagnetic storm not only depends on the interplanetary-magnetospheric coupling parameter VB_z but also on the duration of B_z. The higher the value of $VB_z\ T_{B_z}$ the higher will be the strength of the geomagnetic storm (Srivastava & Venkatakrishnan 2004; Srivastava 2005a). Amongst the solar variables, the initial speed of a CME can be considered as the most reliable predictor of the strength of the associated geomagnetic storm. The initial speed not only influences the ram pressure but is also responsible for strengthening the interplanetary-magnetospheric coupling parameter VB_Z.

Using these results as inputs, a simple logistic regression model has been developed. This model successfully predicts the occurrence of intense geomagnetic storms, although it is only moderately successful in predicting super-intense storms.

The results also indicate that as compared to the solar variables, the interplanetary variables are better predictors of the occurrence of geomagnetic storms, which is possibly due to the fact that the relevant solar variables for the prediction of the geomagnetic storms have not been identified as yet. Therefore, one needs to identify key solar variables that can be effectively used to predict the occurrence of a geomagnetic storm. A key solar variable may be the solar surface magnetic field, and its orientation, which can influence the interplanetary magnetic field and its orientation. Further, some of properties of CMEs may get modified due to interaction with the ambient medium as they propagate through the interplanetary medium, which may also affect the predictive capability of the model. These factors will be taken into account for improving the predictive model, in future.

Acknowledgements

The author acknowledges the LASCO/EIT and CELIAS operational teams on SOHO and the ACE team for providing data used in this paper. The author is also thankful to the referee for suggestions to improve the manuscript.

References

Burton, R. K., McPherron, R. L., Russell, C. T. 1975, *J. Geophys. Res.*, **80**, 4204–4214.

Gopalswamy, N., Yashiro, S., Liu, Y., Michalek, G., Vourlidas, A., Kaiser, M. L., Howard, R. A. 2005a, *J. Geophys. Res*, **110(A9)**, doi10.1029/2004JA010958.

Gopalswamy, N., Yashiro, S., Michalek, G., Xie, H., Lepping, R. P., Howard, R. A. 2005b, *Geophys. Res. Lett.*, **32(12)**, doi10.1029/2004GL021639.

McPherron, R., Siscoe, G., Arge, N. 2004, *IEEE transactions on Plasma Science*, **32**, 1425–1438.

Schwenn, R., Dal Lago, A., Huttunen, E., Gonzalez, W. D. 2005, *Ann. Geophys.*, **23**, 1033–1059.

Srivastava, N., Venkatakrishnan, P. 2002, *Geophys. Res. Lett.*, **29**, 1287–1290.

Srivastava, N., Venkatakrishnan, P. 2004, *J. Geophys. Res.*, **109**, A010103, doi10.1029/2003JA010175.

Srivastava, N. 2005a, *Ann. Geophys.*, **23(9)** 2989–2995.

Srivastava, N. 2005b, *Ann. Geophys.*, **23(9)** 2969–2974.